STIMULUS PREFERENCES OF INDIVIDUALS WITH DEVELOPMENTAL DISABILITIES: DIFFERENTIAL OUTCOMES BASED ON ARRAY COMPOSITION, MEASUREMENT METHOD AND SCHEDULE REQUIREMENTS

By

ISER GUILLERMO DELEON

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1997

ACKNOWLEDGMENTS

I wish to express my gratitude to all who have helped make this dissertation possible. Thanks go to all the current and former graduates students at the Florida Center on Self-Injury who contributed to planning and conducting these studies: Juliet Burke, Han-Leong Goh, SungWoo Kahng, Dorothea Lerman, Jana Lindberg, Eileen Roscoe, Bridget Shore, Michele Wallace, and April Worsdell. I would also like to acknowledge the support and assistance of those who served on my committee: Marc Branch, Mary Kay Dykes, Timothy Hackenberg, and Mark Lewis. A special thanks go out to Brian Iwata, my committee chair and advisor, for his invaluable guidance during the past four years and his patience in helping me through the program quickly.

I would further like to express gratitude to my mother, Marta DeLeon, for providing so much love and encouragement throughout my graduate career and to the memory of my father, Iser DeLeon, for providing the incentive, even in his absence, to complete my education.

Finally, I wish to extend my gratitute to Valerie DeLeon for everything she has done in helping me through these past four years. Without her unconditional love and support, none of this would have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

STIMULUS PREFERENCES OF INDIVIDUALS WITH DEVELOPMENTAL DISABILITIES: DIFFERENTIAL OUTCOMES BASED ON ARRAY COMPOSITION, MEASUREMENT METHOD AND SCHEDULE REQUIREMENTS

By

Iser Guillermo DeLeon

December 1997

Chairman: Brian A. Iwata
Major Department: Psychology

The identification of effective reinforcers is central to the habilitation of individuals with developmental disabilities. However, given that such individuals frequently have limited communication skills, they are often unable to express preferences among stimuli, making the identification of effective reinforcers difficult. In response, a number of methods have been developed for the rapid prediction of stimulus preferences. These methods typically involve presenting a variety of stimuli to individuals and measuring the percentage of trials in which a particular stimulus is selected when available or the duration of engagement with manipulable stimuli. Stimuli that produce the highest selection percentages or durations of engagement are predicted to be the most effective reinforcers.

Three experiments were conducted demonstrating that the results of preference assessments can be influenced by variables that are not typically considered when conducting the assessments. Experiment 1 demonstrated that the presence of several food

items can mask preferences for effective nonfood reinforcers when both are included in the same selection array. Experiment 2 addressed the interpretation of preference assessments that result in uniformly low selection percentages. Whereas this outcome may be typically interpreted as indicating little preference for any of the available stimuli, the results of Experiment 2 suggest that such outcomes may also indicate that several of the available stimuli are highly preferred. Experiment 3 addressed response requirements involved in choosing among available alternatives during preference assessments. The results suggested that under some conditions, preferences established under low response requirements may not accurately predict preferences when the response requirements are increased. These findings are discussed in terms of their implications for the training and treatment of individuals with developmental disabilities and for selecting preference assessment procedures that are closely aligned with the specific outcomes sought.

INTRODUCTION

The identification of effective reinforcers is a critical first step in the habilitation of individuals with developmental disabilities. Both the initial shaping and the subsequent maintenance of adaptive skills in such individuals rely on training techniques based on reinforcement principles. Procedures to eliminate destructive behaviors in this population also often involve the delivery of reinforcers for the absence of such behaviors or for the occurrence of more socially acceptable alternative responses. Unfortunately, stimuli selected for use as reinforcers in typical training and treatment regimens are often identified through informal observation or caregiver opinion (Green et al., 1988; Mason, McGee, Farmer-Dougan, & Risley, 1989). Thus, it is possible that a number of treatment failures result from the application of contingencies in which consequences do not function as reinforcing events (Pace, Ivancic, Edwards, Iwata, & Page, 1985).

Research on the training of individuals with developmental disabilities has illustrated the benefits of identifying multiple sources of reinforcement. Egel (1981), for example, compared the effects of using the same (constant) reinforcers versus different (varied) reinforcers during object identification or discrimination tasks with three individuals with developmental disabilities. During the constant reinforcer condition, a single edible item was delivered for each correct response; during the varied condition, three edible items were randomly rotated following each third correct response. He found that varying the reinforcers resulted in higher levels of correct responding and on-task behavior and suggested that the effects were probably due to satiation to the constant reinforcer. Besides such local effects of constant reinforcer delivery, individual preferences may simply shift over time, so that a single item may have short-lived reinforcing effects (Kazdin, 1984). In either case, it would be beneficial to identify a

number of effective reinforcers for a given individual to maximize the effects of training or treatment.

Unfortunately, the identification of effective reinforcers for individuals with developmental disabilities is difficult for a variety of reasons. Perhaps foremost, such individuals frequently lack the verbal skills to describe their preferences (Ivancic & Bailey, 1996). Another problem may result from lack of exposure to stimuli that might function as reinforcers (Allyon & Azrin, 1968). Finally, individuals may lack the motor skills required to extract sufficient reinforcement from certain items or activities (Singh & Millichamp, 1987).

To address these problems, a number of methods have been developed for the rapid identification of reinforcing stimuli for individuals with developmental disabilities. These methods are used primarily as predictive tools to select stimuli that are subsequently used as reinforcers and have collectively been termed stimulus preference assessments. Research on the development and evaluation of these methods has included definitive tests, termed reinforcer assessments, to determine whether "preferred" stimuli actually function as reinforcers. Preference assessments and reinforcer assessments are said to differ methodologically and conceptually (Piazza, Fisher, Hagopian, Bowman, & Toole, 1996). That is, the aim of preference assessments is to establish a hierarchy of preference for a large pool of stimuli, whereas the aim of reinforcer assessments is to evaluate the reinforcing efficacy of those stimuli identified as preferred. Both types of assessments share a great degree of procedural similarity. Both often involve exposing individuals to simple schedules of reinforcement while testing the effects of the predicted items in maintaining simple responses. Regardless of whether or not they should be considered distinct procedures, in what follows, I shall follow convention and refer to procedures designed to quickly identify preferred items as preference assessments, whereas procedures designed to evaluate the reinforcing effects of the items identified in preference assessments will be termed reinforcer assessments.

Preference Assessment Methods

Methods for identifying stimulus preferences for individuals with developmental disabilities can be broken down into three general classes. These include caregiver rankings (e.g., Cautela & Kastenbaum, 1967; Windsor, Piche, & Locke, 1994), methods based upon participant approach to or selection of available stimuli (e.g., DeLeon & Iwata, 1996; Fisher et al., 1992; Pace et al., 1985), and methods based on the duration of time that a participant spends manipulating an item when available (e.g., Piazza, Fisher, Hanley, Hilker, & Derby, 1996; Ringdahl, Vollmer, Marcus, & Roane, 1997).

Caregiver opinion. Methods based on caregiver opinion typically involve presenting caregivers with a list of items or activities and asking the caregivers to rate the items according to how reinforcing the items might be to the individual in question. For example, Green et al. (1988) asked caregivers to rate the preferences of individuals with mental retardation for 12 stimuli using a 5-point Likert-type scale that ranged from "least preferred" to "most preferred." Others (e.g., Windsor et al., 1994) have asked caregivers simply to rank order a standard set of stimuli along the dimension of preference. A more structured example is the Reinforcer Assessment for Individuals with Severe Disabilities (RAISD) developed by Fisher, Piazza, Bowman, and Amari (1996). The RAISD is a 10-item questionnaire designed to systematically gather information about potential reinforcers. Responses to the questionnaire are used to determine which items to include in a more rigorous preference assessment. Informants are asked to list different types of sensory reinforcers, edible items, social reinforcers, etc., that the person might enjoy. Afterward, the informants are asked to rank items (only those that could easily be delivered or withdrawn contingent upon target behaviors) according to preference. The result is a list of 16 items or activities that can then be assessed using experimental methods.

Although caregiver opinion methods may be highly expedient, they also might be the least reliable method of identifying reinforcers. For example, Green et al. (1988)

compared the results of a caregiver opinion survey to those of a single-stimulus approach method (see description below) to identify the preferences of seven individuals with profound mental retardation for 12 stimuli that included edible items, social reinforcers (e.g., hugs, hand claps), items that provided sensory stimulation (e.g., tactile mitt, tape player) and a variety of toys. Caregiver ratings about each item were averaged, rank-ordered, and compared to the rank orders based on the approach assessment. Spearman rank order correlations were then calculated for individual sets of rankings. Coefficients ranged between -.33 and .11, indicating that caregiver ratings were not highly correlated with the results of the more systematic assessment. In a subsequent reinforcer assessment, the authors recorded the mean prompt level (ranging from verbal instruction to full physical prompt) required to get the participants to complete a task given delivery of a selected stimulus contingent upon correct responding. The results indicated that the approach-based method more accurately predicted items that would reduce the required prompt level relative to a no-reinforcement baseline. That is, stimuli that were highly ranked in the approach-based method generally reduced the prompt level required for task completion. By contrast, stimuli that resulted in high caregiver rankings did not function as reinforcers unless they were stimuli that also produced high approach-based rankings.

Others have also noted poor correspondence between caregiver rankings and more systematic assessments. For example, Windsor et al. (1994) observed that caregiver rankings for six food items based on presumed preference by eight individuals with severe-profound disabilities produced Kendall rank-order correlation coefficients of 0.39 and 0.37 when compared with results of two approach-based methods. Furthermore, coefficients of concordance, designed to measure the consistency of preference rankings across caregivers, produced a mean coefficient of 0.45, with five of the eight comparisons producing coefficients at or below 0.32. Thus, in addition to producing preference rankings that are highly discrepant with those produced by more systematic assessments, caregiver opinions regarding client preferences vary greatly across caregivers.

Approach-based assessments. A second general method for assessing preference can be loosely characterized as approach methods. These methods typically involve placement of items directly in front of the individuals and measuring the number of trials on which an item is approached when available. Predictions about reinforcer efficacy are then based on relative percentages of trials approached (e.g., Fisher et al., 1992; Paclawskyj & Vollmer, 1995; Windsor et al., 1994). The prototype approach method was described by Pace et al. (1985). These authors presented 16 stimuli to each of six participants. The selection array included food and other items that provided various forms of sensory stimulation (e.g., flowers, tape-recorded songs, heat pad). Additionally, various forms of social interaction were made available (e.g., claps, hugs) by associating each with a physical sign of its availability (e.g., for hugs, the therapist leaned toward the participant with outstretched arms). Four stimuli were presented per session and each stimulus was presented five times per session. Thus, each item was presented a total of ten times over the course of eight sessions. The primary dependent measure was the percentage of times a stimulus was approached (defined as moving toward the stimulus with the body or hand within 5 s of its presentation) when it was available.

The authors then assessed the validity of predictions about reinforcer efficacy during subsequent reinforcer assessments. Items selected on 80% or more trials during the preference assessment were regarded as preferred, whereas those selected on less than 50% were regarded as non-preferred. For each subject, a simple response was selected (e.g., reaching, looking in a particular direction, touching the therapist's hands, etc.) and the subjects were given ten prompts per session to emit the response. During baseline, no consequences were provided for correct responding. During subsequent sessions, either a preferred or non-preferred item was delivered contingent upon each correct response. Comparison of percent correct responding during these reinforcer assessment sessions indicated that those items selected during 80% or more of the preference assessment trials

resulted in higher percentages of correct responding than did items selected in less than 80% of trials during the preference assessment.

Other studies have since replicated the utility of the single-stimulus approach method in identifying effective reinforcers (e.g., Green et al., 1988; Smith, Iwata, & Shore, 1995). However, others researchers noted that stimuli selected on a high percentage of trials using this method have not always functioned as effective reinforcers (Green, Reid, Canipe, & Gardner, 1991; Ivancic & Bailey, 1996). This observations has led to suggestions that single-stimulus approach methods may be prone to overestimation of reinforcer value in some cases (Fisher et al., 1992). Overestimates of reinforcer value may come about because participants are likely to approach any stimulus that is placed in front of them. As such, even items that do not function effectively as reinforcers may get selected on a high percentage of trials.

In response to this potential limitation of the procedure described by Pace et al. (1985), Fisher et al. (1992) developed and evaluated an approach method in which items are presented in pairs rather than individually. These authors suggested that such a procedure may be more accurate because it involves direct comparisons between concurrently available stimuli. In their procedure, 16 stimuli were presented in pairs, one pair per trial, until each item had been paired with each other item. As before, estimates of preference were based on the percentage of trials an item was chosen when available. Preference results using this assessment method were compared to single-stimulus assessment results, using 80% approach as the criterion for high preference in both cases. Across four participants, the single-stimulus assessment resulted in 36 stimuli getting selected on 80% or more of the trials they were available. By contrast, only 9 stimuli met this criterion using the paired-choice assessment. To compare the predictive accuracy of the two methods, concurrent variable-interval reinforcement schedules were arranged in which participants received a stimulus that produced approach percentages above 80% during both assessment methods for one response and received a stimulus that was again

ranked above 80% in the single-stimulus assessment, but below 60% during the paired-choice assessment, for another response. The results indicated that all four participants allocated more responding toward the option that resulted in delivery of the stimulus that was highly ranked in both assessments. As such, the authors concluded that the paired-choice procedure resulted in greater differentiation of preference and more accurately predicted subsequent reinforcement effects.

Variations of the paired-choice assessment have since been used successfully in a variety of contexts. For example, Belfiore, Lee, Vargas, and Skinner (1997) used the procedure to assess preference for math problems. The authors presented two students with a choice between single- or triple-digit multiplication problems on separate, concurrently presented worksheets. The number of digits required to complete each sheet was held constant, so that the problems differed only in the number of digits required to complete each problem. The assessment was conducted over 10 sessions. Both students showed a preference for the single-digit problems. Single-digit problems were then used successfully in a momentum-type intervention to increase completion of multiple-digit problems.

A paired-choice procedure has also been used to enhance the effects of "enriched environment" approaches to the treatment of destructive behavior. Vollmer, Marcus, and LeBlanc (1994) found that when high preference toys, as identified through a paired-choice assessment, were used in an enriched environment treatment for individuals whose behavior seemed insensitive to social contingencies, subjects displayed less inappropriate behavior relative to an enriched environment incorporating less preferred items.

Recent studies on preference assessment have sought to further improve concurrent-choice procedures by making them briefer. Windsor et al. (1994) reasoned that if all available stimuli were presented simultaneously, rather than just two at a time, each stimulus would still be compared to every other stimulus, except that those comparisons could be completed in far fewer presentations and still result in a constant fixed number

of presentations per item. These authors compared a paired-choice procedure and a group presentation method in assessing preference for food items of eight individuals with severe-profound disabilities. The group presentation procedure involved the simultaneous presentation of six food items and prompting participants to select one per trial. After an item was chosen, an identical item was placed back into the stimulus array and the trials was repeated. Ten such trials were conducted per session, and the results were compared to a paired-choice assessment involving the same sets of stimuli. The analysis revealed similar patterns of selection responses across the two assessment types. The mean rank-order correlation between the rankings produced by each method was .748. Additionally, six of the eight participants chose the same item during the highest percentage of trials during the two assessments. However, the assessments also differed in two important ways. The group presentation method required substantially less time to administer. The mean number of minutes required to complete a group presentation session was 7 min, whereas a mean of 16 min was required to complete a paired-choice session. Alternatively, the paired-choice assessment resulted in greater consistency across administrations (mean Kendall rank-order coefficient of concordance was 0.63 for the paired-choice assessment, and 0.48 for the group presentation method). Additionally, the paired-choice assessment resulted in more distinct rankings. That is, it was difficult to make predictions regarding relative reinforcement effects with the results of the group presentation method because a single preferred item could be chosen during each trial, thereby providing no information about the reinforcing efficacy of the unselected items. By contrast, the same item was not available during each trial in the paired-choice procedure, such that participants made choices among less-preferred items. This, in turn, provided more information about the relative preference among lesser preferred items.

A simple adjustment to the group presentation procedure, however, has since resulted in rankings that are similar in their distinctiveness to the paired-choice procedure during the group presentation format. This adjustment involved changing the group-

presentation format so that selected items were not replaced across successive trials. DeLeon and Iwata (1996) compared such a procedure to both the paired-choice and group presentation assessments and found that the revised group presentation format (termed multiple-stimulus without replacement, or MSWO, by the authors) resulted in the distinct rankings produced by the paired-choice procedure, but in the brief amount of time that it took to conduct the group presentation format.

Duration-based assessments. A final general class of preference assessment is based upon measures of duration of engagement with the selected items. That is, items are made available, and measurements are taken on the amount of time, either in percentage of item engagement during total availability time or as a percent of intervals of item engagement. Preference is then indicated by the relative amounts of item engagement for each item. Although duration has rarely been directly evaluated as a method of preference assessment or compared to approach methods, several studies have used it in the process of selecting reinforcers for other purposes (Quilitch, Christopherson, & Risley, 1977; Mace, McCurdy, & Quigley, 1990).

Like the approach-based assessments, variations of duration assessments have involved both individual and concurrent item presentation. An example of the former, described by Piazza, Fisher, Hanley et al. (1996), additionally demonstrated an advantage of duration-based methods over those based on approach percentages. That is, preference measures based on duration of engagement permit a comparison of preference for presented stimuli relative to reinforcement that might be derived from ongoing behavior, sometimes destructive behavior. Specifically, comparisons can be made of engagement with a stimulus versus engagement in or rates of other behaviors that occur while the person has access to each item. Following an initial selection of items through the use of the RAISD (Fisher et al. 1996), Piazza, Fisher, Hanley et al. presented each of 16 stimuli to one participant and 12 stimuli to a second participant for 30 s each per session. These authors measured the duration of interaction with each item during the time it was

available and averaged the durations across sessions. In addition to recording duration of engagement with the items, the authors recorded frequency of self-injurious behavior during each 30 s. trial. By comparing duration of engagement for each item to the frequency of SIB displayed while each item was available, these authors identified stimuli that resulted in both high preference and high frequencies of SIB. These items were subsequently shown to be effective reinforcers during reinforcer assessments. However, when subsequently used in differential reinforcement schedules for the treatment of SIB, the same stimuli actually increased SIB rates above those observed during baseline. As such, by combining duration measures of preference with measures of inappropriate behaviors, the authors were able to predict negative side effects for stimuli that might otherwise be incorporated in treatment contingencies simply because of their reinforcing potential. Shore, Iwata, DeLeon, Kahng, and Smith (1997) have similarly shown that the simultaneous measurement of single-item engagement and inappropriate behavior can lead to useful predictions about the use of reinforcers in the treatment of SIB.

Ringdahl, Vollmer, Marcus, and Roane (1997) recently demonstrated that concurrent presentation of items can also result in accurate predictions about preference. These authors made items simultaneously available during 10 min assessment sessions during which both item engagement and frequency of SIB were recorded for three individuals with developmental disabilities. Items that resulted in high levels of engagement and low levels of SIB were subsequently used in an enriched environment treatment for SIB. Based on the low levels of SIB observed during the enriched environment conditions, the authors concluded that this free-operant preference assessment was useful in identifying items that could maximize treatment effects.

Measures of engagement duration may have additional benefits beyond those described above. For example, functional analyses of destructive behavior indicate that some proportion of these behaviors are maintained by access to tangible stimuli. The

single-stimulus approach, paired-choice, and MSWO procedures all involve the removal of potentially reinforcing items, which might occasion destructive behavior in individuals for whom such behavior is maintained by access to tangible reinforcers. Duration-based assessment, when conducted in the manner described by Ringdahl et al. (1997), at most involves only a single instance of forcible removal of a preferred item (i.e., at the end of the 10-min session), and is therefore less likely to occasion destructive behavior in such individuals.

However, concurrent presentation of items during duration measurement as conducted by Ringdahl et al. (1997) may also suffer the same sort of limitations posed by the group presentation method described by Windsor et al. (1994). That is, given that all items were concurrently available with no attempts to induce selection of lesser-preferred items, individuals might manipulate a single item during the entire assessment, providing little information about the reinforcing efficacy of the other items.

Prediction Errors

Thus far, it has been observed that preference assessments have been generally effective in identifying reinforcing stimuli for individuals with developmental disabilities. However, few studies to date have evaluated the sorts of preferences that have actually been made using these methods. That is, results are generally put to use without careful consideration of the items that have been selected, whether alternative interpretations are possible based on results of assessments, and the generality of the results across conditions that differ from those under which predictions are made. In the following series of investigations, the first experiment examined how the composition of stimulus arrays during assessment could influence results. Specifically, the extent to which a general preference for food items might mask the reinforcing potential of nonfood items was explored. A second study attempted to determine how best to interpret assessments that result in uniformly low approach percentages during MSWO assessments. Finally, recent research has suggested that results of assessments conducted under low schedule

requirements may not accurately predict reinforcement effects when effort is increased. Thus, the third study is an attempt to replicate this research and begin to determine the conditions under which this sort of prediction error might occur. Given the varied nature of these investigations, each study will be introduced individually.

EXPERIMENT I: DISPLACEMENT OF LEISURE REINFORCERS BY FOOD DURING PREFERENCE ASSESSMENTS

Casual inspection of some of the studies discussed above suggests that food items may be selected disproportionately more often than nonfood items. For example, in the study by Pace et al. (1985), food items were selected more than or as often as the nonfood items by four of the six participants. Similarly, Smith et al. (1995) presented 14 items to four participants using the Pace et al. procedure. Only one food item was available, and it was chosen on 100% of the trials by all four participants.

Although the purpose of these studies was not to examine relative preference between classes of reinforcers, the results suggested that individuals with severe disabilities may have a general preference for food items relative to nonfood items. If so, one determinant of choices in the context of preference assessments may be the pool of items from which selections are made, or the extent to which food items are mixed in with nonfood items. If food is generally preferred, then the reinforcing potential of nonfood items may be obscured during preference assessments simply because nonfood items do not compete well with food items.

In the first experiment, I sought to determine the extent to which (a) individuals with developmental disabilities preferred food items over leisure items, and (b) whether leisure items that are "displaced" by food items nevertheless function effectively as reinforcers.

Method

Participants, Settings, and Materials

Fourteen individuals participated in the study. All but one were diagnosed with profound mental retardation; the exception was Eliza, who was diagnosed with moderate

mental retardation. All lived in a public residential facility for persons with developmental disabilities and had been referred to a day clinic for the treatment of self-injurious behavior. Sessions were conducted in one of the rooms of the clinic. Two of the subjects had visual impairments. Chuck had myopia but wore no corrective lenses because his vision permitted him to function adequately. Eddie was blind in his right eye. Experimenters were graduate and undergraduate students working at the clinic.

For each participant, 14 stimuli were initially selected for assessment. Seven stimuli were selected from a list that included sweet, salty, bitter, and sour foods. The other seven stimuli were selected from a list of leisure items that included stimuli primarily affecting visual, auditory, tactile, or olfactory sensory modalities or were selected based on staff opinions about participants' preferred leisure items.

Procedures and Experimental Design

Stimulus preference assessments. All participants were exposed to three multiple stimulus without replacement (MSWO) preference assessments (DeLeon & Iwata, 1996), each conducted over five sessions. Seven items were presented during each assessment. All sessions began with the participant and an experimenter seated in adjacent chairs at a table. After placing all of the items in an array in front of the participant, the experimenter asked the participant to choose one item. Following each selection, the participant was allowed to consume a single food selection (e.g., one pretzel) or to have access to selected leisure items for 30 s. Selected items were not replaced in the array. The session continued until all items had been selected or a 30-s period with no selections had elapsed. The primary measure was a percentage reflecting the number of times an item was selected divided by the number of times that item was available during the five sessions, multiplied by 100%. Items were subsequently ranked according to selection percentages.

In the food assessment, the stimulus array consisted entirely of small edible items and small quantities of drinks. In the leisure assessment, the array consisted of nonfood

items only. Some individuals were exposed to the food assessment first, whereas others were exposed to the leisure assessment first. Following the completion of the separate food and leisure assessments, the top-ranked stimuli from both assessments were combined into a third array composed of both food and leisure items. The combined array typically consisted of the top three items from the first assessment and the top four items from the second assessment unless fewer than the requisite number of items had been selected during the initial assessments. In such cases, the combined array consisted of all of the items selected from one assessment (the array that failed to reach the three- or four-item criterion) and as many items from the other assessment as needed to complete a seven-item combined array.

Reinforcer assessment. With two participants, Sheila and Alex, an attempt was made to determine if a leisure item (Connect 4® toy in both cases) that was ranked below at least three food items in the combined (third) assessment would function effectively as a reinforcer for an adaptive response. For both subjects, this item was ranked fourth in the combined assessment, but ranked first in the leisure assessment. In the combined assessment, Sheila selected this toy on 23.8% of the trials it was available, whereas the selection percentage for Alex was 22.7%.

Sheila engaged in a high frequency of hand mouthing; therefore, drying her hands, defined as taking a towel from the experimenter and wiping either hand with it, was selected as her target response. Alex's target response, towel folding, was selected from those listed as habilitation goals in his service plan. For the purpose of the present study, the response involved only the final step of a task analysis and was defined as taking a folded towel from the table and placing it into a basket.

All sessions lasted 5 minutes, corrected for access to the stimuli. That is, the experimenter kept track of total session time on a hand-held timer, stopped the time upon delivery the leisure item , and restarted the timer upon the removal of the item. The experimenter verbally prompted the participant to emit the target response at the

beginning of each session (e.g., "Sheila, wipe your hands.") and at 1-min intervals throughout the session. During baseline, each occurrence of the target response resulted in praise from the experimenter (e.g., "Thanks for wiping your hands, Sheila"). During the reinforcement condition, each occurrence of the response resulted in 30 s access to the Connect 4®. Baseline (A) and reinforcement (B) conditions were presented in a reversal design (ABAB), and the dependent measure of interest was the frequency of responses (per minute) during each condition.

Data Collection and Interobserver Agreement

During the stimulus preference assessments, trained graduate and undergraduate student observers recorded the order of item selection using data sheets specifically designed for this purpose. During 36.2% of the assessment sessions, a second observer recorded the order of selections for purposes of assessing interobserver agreement. When comparing observers' records, an agreement was scored if both observers recorded the same order of selection for each item. Interobserver agreement was calculated by dividing the number of selections on which observers agreed by the total number of selections and multiplying by 100%. All agreement scores were 100%.

During the reinforcer assessments, the same observers collected data using a hand-held computer (Assistant, Model A102). Observers recorded the occurrence of prompts, adaptive responses, and reinforcer deliveries. Data were collected by a second observer during 32.8% of the sessions. When comparing observers' records, session time was divided into 10-s intervals. Interobserver agreement was calculated on an interval-by-interval basis by dividing the smaller number of recorded events by the larger number. These quotients were then summed across intervals, divided by the total number of intervals in the session, and multiplied by 100%. Mean interobserver agreement was 97.3% (range, 80.0% to 100%) for prompts, 95.2% (range, 78.3% to 100%) for adaptive responses, and 95.1% (range, 88.3% to 100%) for reinforcer delivery.

Results

Table 1 shows the selection rankings and means obtained for the food and leisure items during the combined assessment for each participant. For 12 of 14 participants (85.7%) , the most frequently selected (i.e., highest ranked) item was a food. For 11 of 14 participants (78.6%), the two most frequently selected items were foods. For 9 of the 13 participants (69.2%) whose combined arrays included at least three food items, the three most frequently selected items were foods. Finally, for 8 of the 14 participants (57.1%), the lowest ranked food was selected more often than the highest ranked leisure item.

Table 1
Rankings of Food and Leisure Items for the Combined Stimulus Preference Assessment for Each Participant

Participant	Food Item Ranks	Leisure Item Ranks
Chuck	1, 2, 3, 4, 5	6, 7
Jim	1, 2, 3, 4	5, 6, 7
Alex	1, 2, 3, 5	4, 6, 7
Sheila	1, 2, 3	4, 5, 6, 7
Robbie	1, 2, 3	4, 5, 6, 7
Dina	1, 2, 3	4, 5, 6, 7
Janet	1, 2, 3	4, 5, 6, 7
Rod	1, 2, 3	4, 5.5, 5.5, 7
Charlene	1, 2, 3.5, 3.5	5, 6, 7
Reggie	1.5, 1.5, 4	3, 5, 6, 7
Rudy	1, 3, 4	2, 5, 6, 7
Carly	1, 2, 6	3, 4, 5, 7
Eliza	2, 4, 5, 7	1, 3, 6
Eddie	2, 3.5, 6, 6, 6	1, 3.5

Figure 1 shows the selection percentages for leisure items during the leisure assessment and during the combined assessment. Of the 48 leisure items tested in the combined arrays, 45 (93.7%) were selected on a lower percentage of trials during the combined assessment relative to the leisure assessment. Only the fourth ranked item for Carly, the second ranked item for Eliza, and the highest ranked item for Eddie were selected more often during the combined assessment. Additionally, 45 of the 48 leisure items (93.7%) produced lower rankings during the combined assessment relative to the leisure assessment. These included the same items described above for Carly and Eddie, and the highest ranked item for Eliza.

Spearman rank-order correlations were calculated between rankings for leisure items from the leisure assessment and rankings for the same items when included in the combined assessment (not including the foods) to assess participants' consistency across the two assessments. For example, if items ranked first, second, third, and fourth in the leisure assessment were ranked fourth, fifth, sixth, and seventh, respectively, in the combined assessment by a given participant, then ranking consistency was retained even though the leisure items were displaced by food. This analysis yielded a mean rank-order correlation of .692 (range, -.389 to 1.00). With the exception of Rod, all correlations were positive and higher than .50 (the mean correlation excluding Rod was .776).

Figure 2 shows the selection percentages for foods during the food assessment and during the combined assessment. Only 23 of the 50 foods (46.0%) were selected on a lower percentage of trials during the combined assessment relative to the food assessment. In terms of rank comparisons, 20 of the 50 food items (40.0%) produced lower rankings during the combined assessment relative to the food assessment, whereas 16 (32.0%) resulted in higher rankings. The remaining 14 food items (28.0%) retained the same ranking across assessments.

Spearman rank-order correlations across food and combined assessments for the food items yielded a mean coefficient of .335 (range, -.949 to 1.00), indicating that,

Figure 1. Selection percentages for leisure items during the leisure assessment and during the combined assessment (Experiment 1).

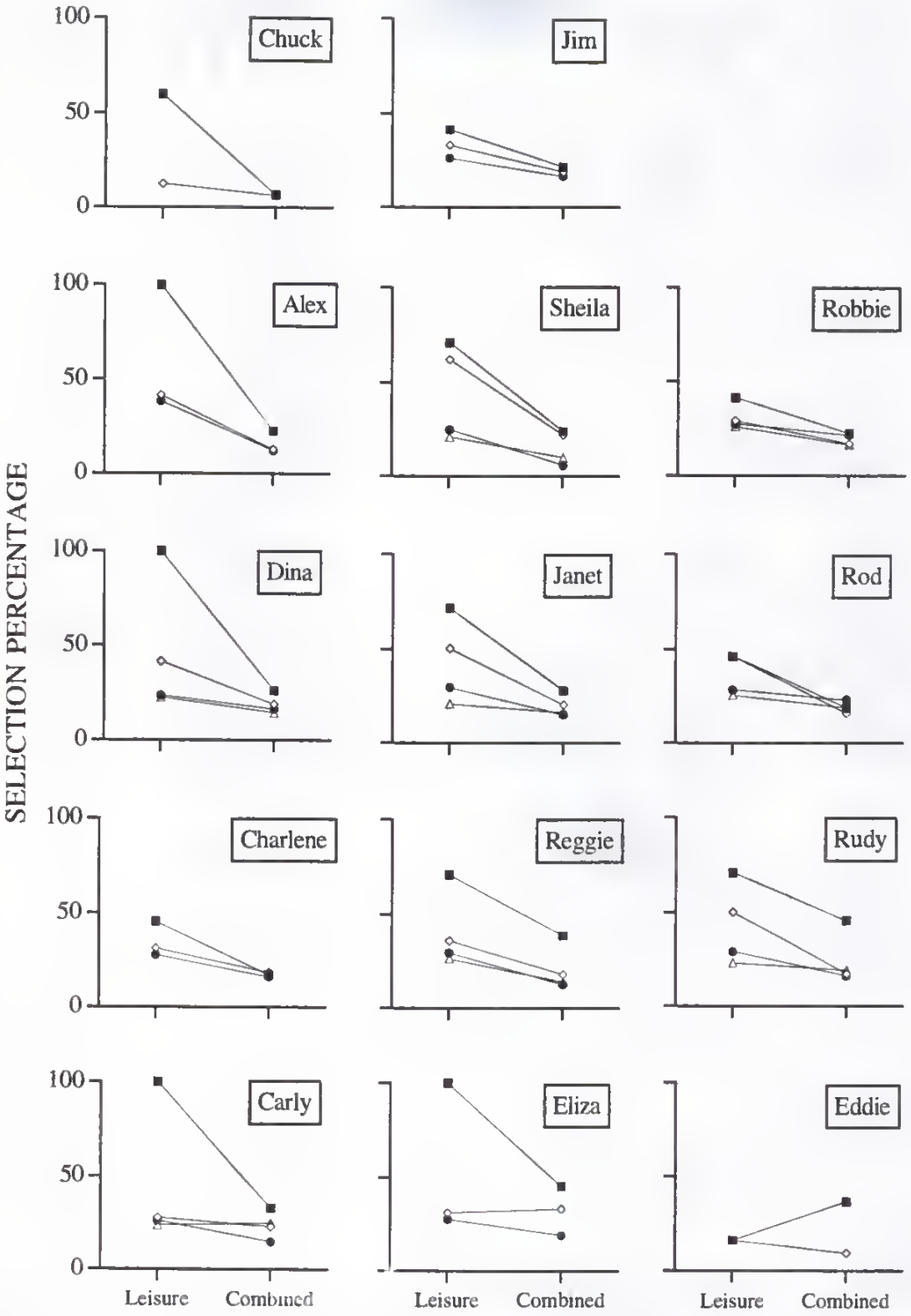
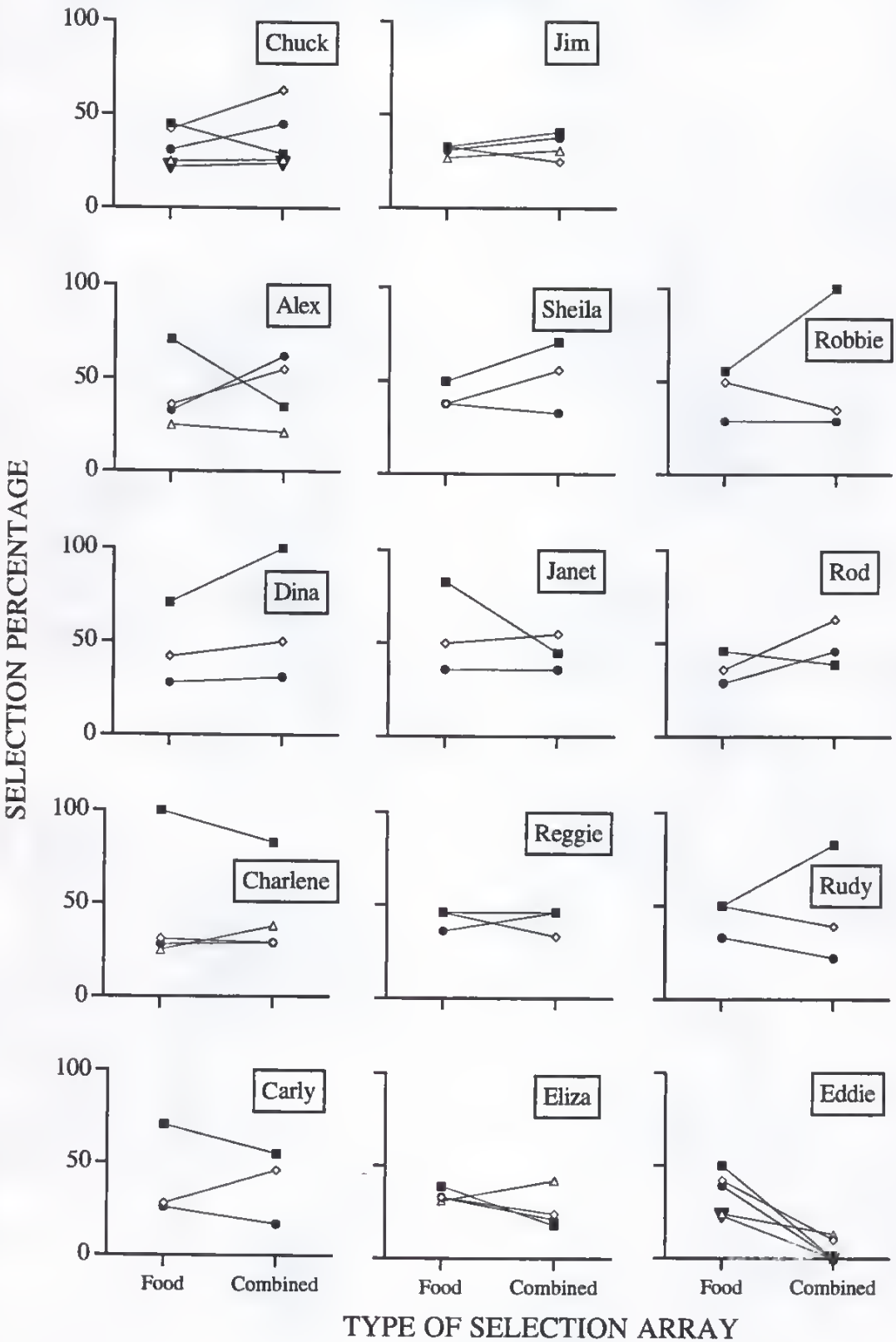

Chuck
Jim
Alex
Sheila
Robbie
Dina
Janet
Rod
Charlene
Reggie
Rudy
Carly
Eliza
Eddie
100
50
0
Leisure
Combined
SELECTION PERCENTAGE
TYPE OF SELECTION ARRAY

Figure 2. Selection percentages for food items during the food assessment and during the combined assessment (Experiment 1).


Chuck
Jim
Alex
Sheila
Robbie
Dina
Janet
Rod
Charlene
Reggie
Rudy
Carly
Eliza
Eddie
100
50
0
Food
Combined
SELECTION PERCENTAGE
TYPE OF SELECTION ARRAY

relative to the leisure items, the rank ordering of foods was less consistent across the two assessments. However, this also suggests that the displacement found among food items, whether upward or downward in terms of rank order, was a function of displacement by other food items rather than downward displacement of foods by leisure items.

Figure 3 shows the results of the reinforcer assessments for both Sheila and Alex. During baseline, Sheila (top panel) displayed a low, steady rate of hand drying. During the first reinforcement condition (FR 1 Connect 4), her responding was more variable but showed a substantial increase over that observed during baseline. Sheila's responding decreased during the return to baseline, and increased again during the final reinforcement condition. Alex's rate of placing towels in the basket (bottom panel) showed a slight and gradual increasing trend during the initial baseline. An accelerated increase in responding was observed during the first reinforcement condition, followed by an immediate decrease during the return to baseline, and another increase during the final reinforcement condition.

Discussion

After obtaining distinct rankings for stimuli during preference assessments when food and leisure items were presented separately to 14 individuals with developmental disabilities, it was observed that a large majority of participants later showed a strong preference for food when food and leisure items were combined in the same assessment. That is, during the combined assessment, food items readily displaced leisure items that were highly preferred in the absence of food. As a result, selection of leisure items was generally low during the combined assessment. The fact that displaced leisure items did serve as reinforcers for two participants' adaptive responding suggests that when food and leisure items are combined in stimulus preference assessments, the resulting outcomes for leisure items may represent false negatives. That is, if only the combined assessment were conducted, based on their relatively low ranks and selection percentages, a likely prediction would have been that leisure items were not very effective reinforcers. The

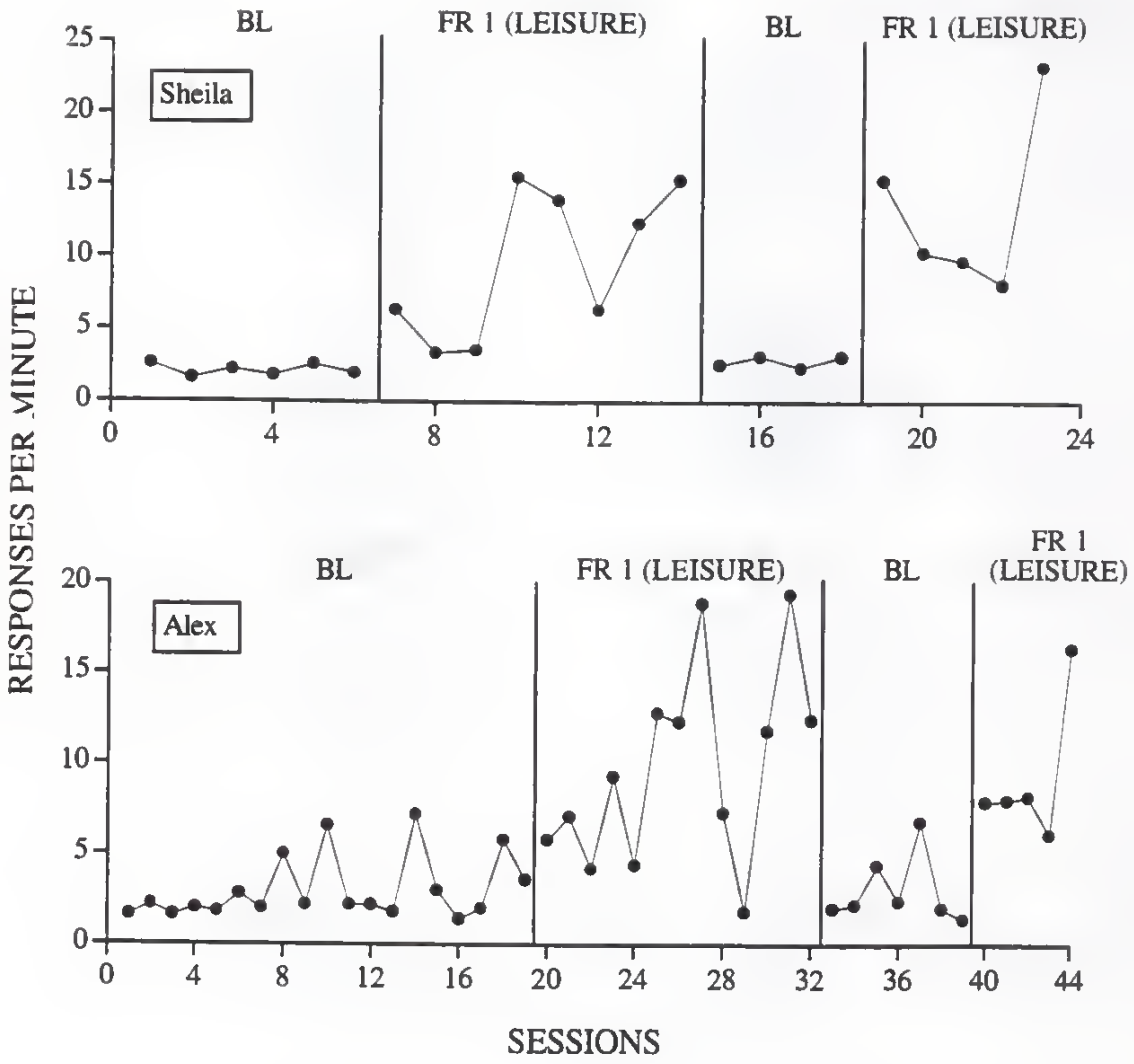


Figure 3. Responses per minute of adaptive behavior emitted by Sheila (top panel) and Alex (bottom panel) during baseline and FR1 delivery of leisure items (Experiment 1).

extent to which items ranked low in preference assessments function effectively as reinforcers is largely unknown. Piazza, Fisher, Hagopian et al. (1996) demonstrated that items from the middle of a ranking maintained higher levels of responding than did items from the bottom of a ranking. However, common usage dictates that only the top few items are usually used in training or treatment. If so, the present results suggest that some effective leisure reinforcers would have been masked by their combination with food in preference arrays.

To the extent that these data are generalizable, the displacement effect observed here would influence the results of any preference assessment in which foods and leisure items are assessed together and in which stimuli are presented more than one at a time. For example, DeLeon and Iwata (1996) reported that food items tended to be the higher ranked items when preference was assessed using both the multiple stimulus format described in this study the paired stimulus format (Fisher et al., 1992), in which stimuli are presented two at a time. The only format for which selective preference for one class of reinforcers would not affect results obtained for other classes of reinforcers is one in which stimuli are presented individually (Pace et al., 1985).

It is possible that the displacement of leisure items observed in this study was not so much an effect related to the presence of food but, rather, to the presence of several highly preferred items that just happened to be food. In other words, displacement might have occurred even if the other items were not food, simply by virtue of having highly preferred items compared with less preferred items. However, a selection of randomly chosen, albeit highly preferred, leisure items would not have produced results as consistent as those observed here. Some displacement would have occurred, but upward and downward displacement from separate assessments would have been more equally distributed across the two sets of stimuli when they were combined. By contrast, the present results indicated that leisure items were consistently displaced downward.

It is also possible that the observed preference for food may have reflected the effects of establishing operations (Michael, 1982), such that participants' motivation to obtain food was generally, but only temporarily, higher than that to obtain leisure items. Previous research has shown that food functions more effectively as reinforcement prior to, rather than following, scheduled mealtimes (Corte, Wolf, & Locke, 1971; Vollmer & Iwata, 1991), leading to a speculation that participants in this study may have been somewhat deprived of food during the assessments. Sessions were typically run between 9:00 a.m. and 12:00 p.m. or between 1:00 p.m. and 4:00 p.m., with participants attending for only one shift per day. As such, the morning participants were exposed to the assessments approximately 2-5 hours following breakfast, whereas the afternoon participants were assessed somewhat sooner following lunch. If relative food deprivation contributed to the general preference observed for food, one might expect a larger effect for the morning participants. However, inspection of preferences exhibited by morning versus afternoon participants revealed no orderly differences in preference for food versus leisure items as a function of assessment time. In future studies, relative levels of deprivation or satiation to both food and leisure items might be manipulated systematically to determine the influence of establishing operations on the results of preference assessments. It is certain that some amount of deprivation or satiation will strengthen or weaken reinforcement effects, but whether events such as meals, recreational activities, or work would exert a disruptive influence during the brief exposures to stimuli typically used during preference assessments is unknown.

Results obtained in this study consistently showed that food was a highly preferred event and are consistent with results from an extensive body of research demonstrating that food can be an effective reinforcer for establishing and maintaining a variety of adaptive behaviors, as well as for reducing the frequency of behavior problems. However, as a class of reinforcers, food items may present some disadvantages relative to nonfood items. Rincover, Newsom, Lovaas, and Koegel (1977) pointed out that the

contingent delivery of food does not often occur in naturalistic settings. There are also some concerns over the motivational properties of food reinforcers. Rincover and Newsom (1985) observed that participants tended to satiate more quickly to food reinforcers than to sensory reinforcers and concluded that food may function inconsistently as a reinforcer due to ethical and legal standards related to food deprivation. By contrast, fewer concerns seem to be expressed regarding the withholding of leisure items, and research has indicated that their reinforcing effects can be enhanced through the manipulation of establishing operations (Vollmer & Iwata, 1991). Additional problems may stem from the fact that the delivery of food reinforcement may not occasion very much caregiver-client interaction (Rincover & Newsom) and that, outside the context of full meals, food consumption does not require very much behavior on the part of an individual in order to extract reinforcement once the item has been earned. By contrast, engagement with many leisure reinforcers requires participation in social activities with others or, alternatively, extended durations or sequences of behavior involving item manipulation. These factors may be most critical in the suppression of behavior problems due to behavioral competition; that is, engaging in appropriate social interaction or solitary leisure activity occupies time that might otherwise be spent engaging in the problem behavior. Finally, Rincover and Newsom pointed out that extensive use of food reinforcers may expose individuals to health and dental risks.

For these reasons, effective nonfood reinforcers may be especially important in the training and treatment of individuals with developmental disabilities, and the present results indicate that special precautions may need to be taken when attempting to identify such reinforcers. Specifically, single-stimulus presentation may be required initially, or, if stimuli are presented in pairs or groups, nonfood items may need to be assessed separately. If, under these conditions, little preference is observed for leisure items, then specific training may be required to establish object manipulation as a reinforcing activity per se (Singh & Millichamp, 1987). Perhaps if individuals were better skilled in

extracting appropriate reinforcement from a subset of leisure items, they might show greater preference for leisure items during an assessment, even though they may have had little experience with the specific stimuli being presented. For example, it would be interesting to see if, after training is provided with selected leisure items, relative preference for these or other leisure items increases relative to that observed for food reinforcers.

EXPERIMENT II: COMPARING DURATION AND APPROACH METHODS FOR IDENTIFYING LEISURE ITEM REINFORCERS

As previously described, methods of identifying reinforcers for individuals with developmental disabilities have often relied on the occurrence of an approach response as a predictor of reinforcement effects. As seen in the first experiment, approach measures might present certain problems when leisure item reinforcers are compared with food. Additionally, it was noted in Experiment 1 that some of the participants failed to select any of the leisure item reinforcers on a high percentage of trials, even when these items were presented in the absence of food (i.e., during the leisure only assessments). For example, five of the participants in Experiment 1 (Jim, Robbie, Rod, Charlene, and Eddie) failed to select any of the leisure items during more than 50% of the trials when the items were available in the leisure only assessments.

Failures to produce high selection percentages may be a problem inherent in all methods that involve simultaneous presentation of more than one stimulus (e.g., DeLeon & Iwata, 1996; Fisher et al., 1992; Windsor et al., 1994). Specifically, predictions based on these types of assessments may be difficult to interpret when the results reflect low selection percentages for most items. One interpretation is that low selection percentages may reflect indifference to the available items. That is, none of the items is highly preferred, so the individual may be more-or-less randomly selecting items only because he or she is being prompted to do so. Given the manner in which preference assessments are typically interpreted, it is likely that uniformly low selection percentages would be taken as an indication that none of the items is a particularly effective reinforcer.

However, a second possible interpretation is that the individual had multiple strong preferences such that it was unlikely for the person to consistently select one item over

the others. For example, assume a child finds several kinds of video games to be extremely reinforcing, but has no strong preference for puzzles. If any of the video games was, by itself, compared to a selection of puzzles in a preference assessment, a clear preference might emerge for the video game. However, given a selection array consisting of a variety of video games, the child might sometimes select one or another, depending on the momentary motivation for any particular game at the time of selection. If that motivation changes slightly from selection to selection, a seemingly random distribution of selections might again result in low selection percentages. However, in this case, low percentages might actually reflect that more than one of the choices are potent reinforcers.

One way to determine which of the above interpretations is more accurate is to use single-item approach methods because stimuli are not concurrently available, thereby eliminating the possibility that low selection percentages reflect direct competition among highly preferred items. However, as Fisher et al. (1992) noted, single-item approach measurements may also be misleading because of simple demand characteristics of the assessment arrangement. That is, perhaps participants are highly likely to approach anything that is placed in front of them. Once having approached the item, however, they may not be likely to interact with that item unless it is reinforcing. For example, Paclawskyj and Vollmer (1995) reported that each of 18 items placed in front of three participants was approached on over 70% of trials during a procedure similar to the one reported by Pace et al. (1985), and 13 were approached every time they were available. However, not all of these items were effective in maintaining rates of compliance on adaptive tasks at levels higher than baseline. The authors anecdotally reported that one of these items was thrown to the ground when presented.

Combining single-stimulus presentation with a duration measure of preference is another alternative that may avoid both the problems of artificially low selection percentages presented by multiple-stimulus presentation formats and artificially high

selection percentages presented by the single-stimulus approach method. Like the single-stimulus approach method, duration measures need not directly compare one stimulus against another or several others. Instead, items can be presented individually, producing individual measurements of the percent of time the person engages with the item when it is made available. These percentages can then be compared in making predictions about reinforcer effects.

However, duration measures may be less sensitive to demand characteristics than mere approach. That is, an individual might similarly approach an item when prompted, but might be less likely to maintain contact with that item unless it produces sufficient reinforcement. In this way, measures of duration could help to clear up the interpretation of uniformly low selection percentages. Given a set of stimuli for which no stimulus was selected on a high percentage of trials, in cases in which none of the items is preferred, a duration measure should reflect low duration of engagement measures for all or most of the items. Alternatively, in cases in which many of the items are highly preferred, a duration measure should result in a high percentage of engagement duration for several of the items.

In Experiment 2, sets of items that resulted in uniformly low selection percentages in an MSWO assessment were individually assessed using a duration measure. Subsequently, items for which a discrepancy was found (i.e., low selection percentage, but high engagement duration) were subjected to a single-operant reinforcement assessment to determine if the predictions made by high duration measures were more accurate, in terms of identifying effective reinforcers, than predictions based on low selection percentages.

Method

Subjects, Settings, and Materials

Three of the individuals from Experiment 1 (Robbie, Rod, and Charlene) participated in Experiment 2 . A fourth participant, Max, had also been referred to a day

treatment program for the assessment and treatment of self-injurious behavior. Max was a 34-year-old male diagnosed with profound mental retardation. He could follow very simple directions, but displayed no expressive language. For three of the subjects, all sessions were conducted in one of the rooms of the treatment program. Sessions for Charlene were conducted in the bedroom of her residence. For each participant, seven items were assessed during the preference assessments. These items were arbitrarily selected from a list of activity reinforcers that included items that primarily affected visual, olfactory, auditory, or tactile sensory modalities.

Procedures and Experimental Design

Preference assessments. Two assessments of stimulus preference were conducted with each participant, both involving the same set of seven stimuli. One assessment used the MSWO format (DeLeon & Iwata, 1996) described in Experiment 1. For Robbie, Rod, and Charlene, these are the same leisure item assessments conducted in Experiment 1. The second method used duration of item engagement as the primary dependent variable. The subjects and experimenters were again seated in adjacent chairs at a table. During each trial, the experimenter placed a single item directly in front of the participant for 2 min. No instructions were given. Observers started a timer when the item was first placed in front of the participant. A second timer was started when the participant first made contact with an item and was kept running until the participant placed the item back down or until the end of the 2 min trial. If the participant released the item during the 2-min period and then picked it back up, the second timer was stopped and restarted so that the total number of seconds the individual engaged the item was recorded regardless of how many times it was released and picked up. At the end of the 2-min trial, the experimenter removed the item from in front of the participant or from the participant's hands. After a few seconds break, the next item was presented in the same manner. This procedure continued until each item had been presented for 2 min. Five such sessions

were conducted with each participant, and results were summarized as the percentage of time (of 10 min total) during which the participant manipulated each item.

Reinforcer assessments. Two participants, Charlene and Robbie, were subsequently exposed to reinforcer assessment procedures similar to those used in the first experiment. Three items were used across these two participants. Two of the items (one for each participant) were chosen because they had been selected during less than 50% of trials available during the MSWO assessment, but had also resulted in 50% or greater item engagement during the duration assessments. These items were a kaleidoscope for Charlene and a vibrating massager for Robbie. Charlene selected the kaleidoscope during 21.7% of the trials it was available during the MSWO assessment, but engaged the item for 58.0% of the time it was available during the duration assessment. Robert selected the massager during 25.0% of trials in the MSWO assessment, but engaged the item for 90.0% of the time during the duration assessment. As such, different predictions might be made about the reinforcer potency of these items based on the two assessment methods. That is, given that both items were selected during a low percentage of time during the MSWO assessment, as typically interpreted, results would suggest that neither of the items was a potent reinforcer. However, based on the percentage of time the participants engaged the items, the duration assessment might predict that both items would be reinforcers. To further validate the accuracy of duration measure predictions, the reinforcing effects of a second item was briefly assessed for Robbie. A See and Say toy was selected during 27.8% of trials available during the MSWO assessment and engaged during 39.0% of the time during the duration assessment. Based on these low percentages, both methods would have predicted that the toy was not an effective reinforcer.

The response used during Charlene's reinforcer assessment was hair brushing, defined as picking a hairbrush up from a table in front of her and running the bristles through her hair at least once. Prior to each session, she was seated in a chair in her

bedroom, and a small table was placed directly in front of her. At the beginning of the session, the brush was placed on the table. Robbie's response involved matching coins to groups of similar coins. Robbie was seated at a table that held four small groups of coins, separated into pennies, nickels, dimes, and quarters. At the beginning of a session, a coin was placed in front of Robbie. A correct response was defined as placing or sliding the coin into the pile with matching coins. Following each correct response, a new coin was immediately placed on the table in front of him.

Sessions proceeded in the manner described in Experiment 1. That is, all sessions lasted 5 min, corrected for reinforcement time, and verbal prompts were delivered at the beginning of the session and at the beginning of each minute thereafter (i.e., 5 prompts per session). Charlene was simply prompted to brush her hair, Robbie was prompted to "match the coin with the correct group." During baseline, correct responses resulted in verbal praise; during reinforcer assessment phases, correct responses resulted in 30 s access to the item being tested. As before, the assessment proceeded in an ABAB design, where A = baseline and B = FR 1 delivery of the item. A third condition was added for Robbie after his second FR 1 phase, during which the See and Say toy (the item that was selected on a low percentage of trials and engaged during a low percentage of time) was delivered for correct coin matching on an FR 1 schedule.

Data Collection and Interobserver Agreement

Trained graduate and undergraduate students observers recorded the order of item selection during the MSWO procedure. A second observer recorded selections during 40% of sessions and interobserver agreement was calculated as described in Experiment 1. All agreement scores were 100%.

A second observer measured duration of item engagement during 45.0% of the duration assessment sessions. Interobserver agreement for duration engagement was measured as percent agreement on duration within a 3 sec window. That is, an agreement was scored if the reliability observer's total duration for a given item was within 1 s plus

or minus of the primary observer's duration. The number of agreements across all participants was then divided by the number of agreements plus disagreements and multiplied by 100. Across all observations, the percentage agreement was 88.7%. The largest discrepancy was 5 s and another was 3 s. All other disagreements were within 2 s.

The same data collection and interobserver agreement procedures described for reinforcer assessments in Experiment 1 were used during Experiment 2. A second observer recorded occurrences of prompts, adaptive responses, and reinforcer deliveries during 24.0% of sessions. Mean interobserver agreement was 100% for prompts, 98.0% (range, 85.7% to 100%) for adaptive responses, and 97.1% (range, 85.7% to 100%) for reinforcer delivery.

Results

Figure 4 displays the selection and duration engagement percentages for each item for each subject. Across subjects, only one of the items resulted in a selection percentage of over 50%. Specifically, Max selected the blocks during 55.6% of the trials they were available. Robbie's highest percentage item, a music box, was selected during 41.7% of trials. Rod selected both a music box and a vibrating massager during 45.5% of trials. Finally, Charlene's highest percentage item was again a massager, chosen on 45.5% of the trials it was available. None of the items failed to be selected at least once during the assessments.

By contrast, 23 of the 28 items across subjects were engaged for over half of the total time they were available during the duration assessments. Max engaged six of the seven items for over 50% of the time. The highest percent duration was 98.8% for a puzzle. Robbie engaged five of the items for over 50% of the time; the highest percent duration was obtained for a massager (90.0%). Rod engaged all seven items for over half the available time, with a highest percentage of 98.2% for a magazine and a lowest of 63.5% for crayons and paper. Finally, Charlene held five of seven items for over half the time, with a music box producing the highest percentage (94.7%).

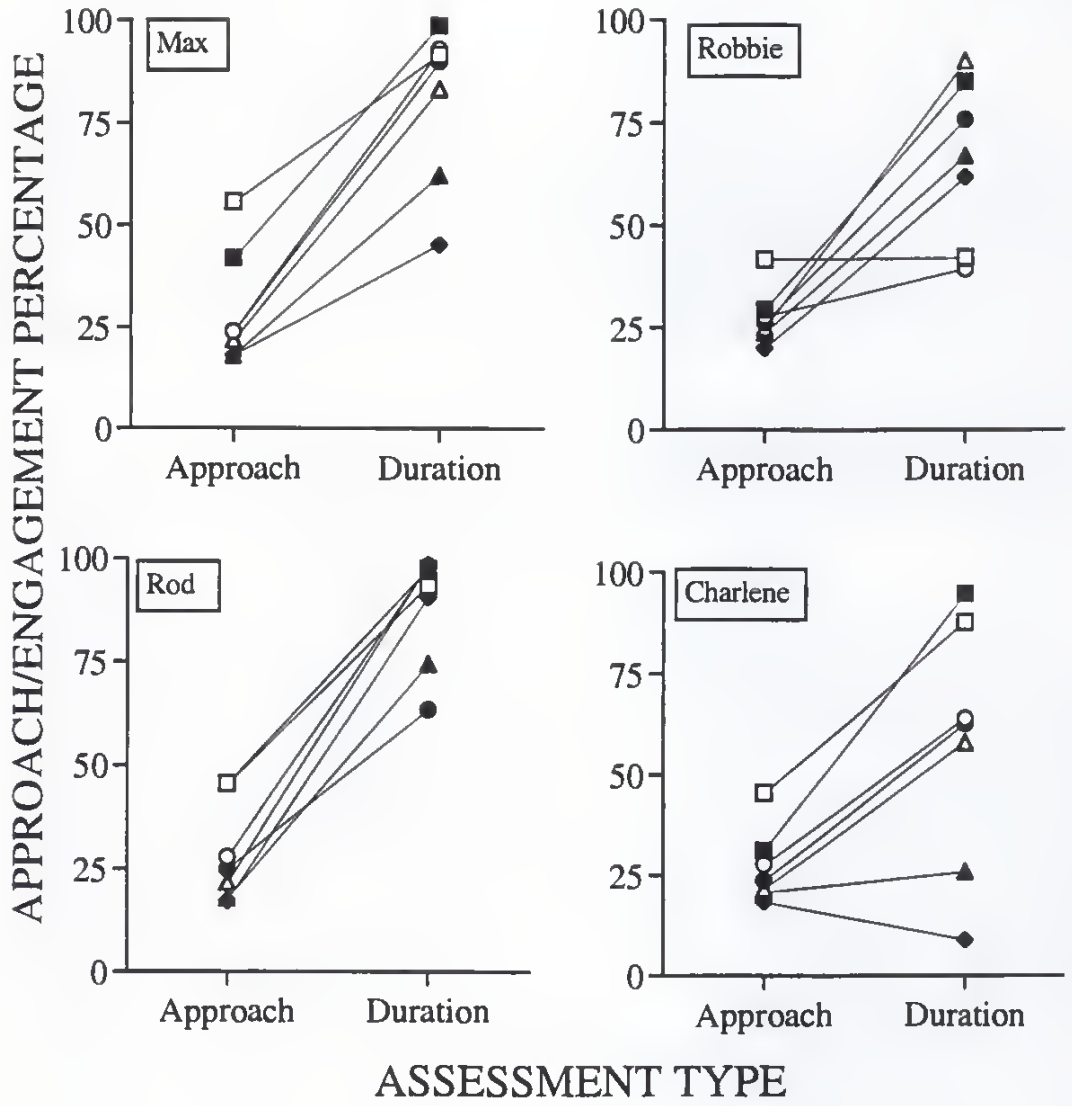


Figure 4. Selection and duration engagement percentages for each item for each subject (Experiment 2).

Figure 5 shows the results of the reinforcer assessments for Charlene and Robbie. During the initial baseline condition, Charlene displayed hair brushing at mean rate of 0.05 responses per min. No responding was observed during the first FR 1 session, but responding increased thereafter during this phase, yielding a phase mean of 2.6 responses per min. A reversal to baseline conditions resulted in a decrease to a mean of 0.15 during this phase, followed by an increase to 5.85 response per min during the final FR 1 condition.

A similar pattern was displayed by Robbie during the assessment of the low-selection/high-duration stimulus. Response rates throughout the assessment were far more variable than those observed for Charlene, but responding was again substantially higher during the FR 1 conditions relative to the baseline conditions. The mean rate during the first baseline condition was 4.1 responses per minute. This increased during the first FR 1 condition to 15.7 responses per min. Reversals to baseline and FR 1 conditions resulted in a decrease to a mean of 7.4 response per min followed by an increase to a mean of 19.3 responses per min. In the final condition, FR 1 delivery of the See and Say toy (a stimulus that produced low selection and duration percentages) failed to produce any increases above the rates observed during baseline. The mean rate for this condition was 5.7 responses per minute.

Discussion

In this study, two methods of predicting reinforcement effects were compared. Sets of items that resulted in uniformly low selection percentages when incorporated into a multiple-stimulus approach preference assessment were subsequently assessed using a single-stimulus presentation method in which duration of item engagement was used as the measure of preference. For all four participants, several of the items that produced selection percentages below 50% produced duration percentages above 50%. This suggested that, as typically interpreted, the approach-based assessment might not have predicted that these items would function as reinforcers, whereas an assessment based on

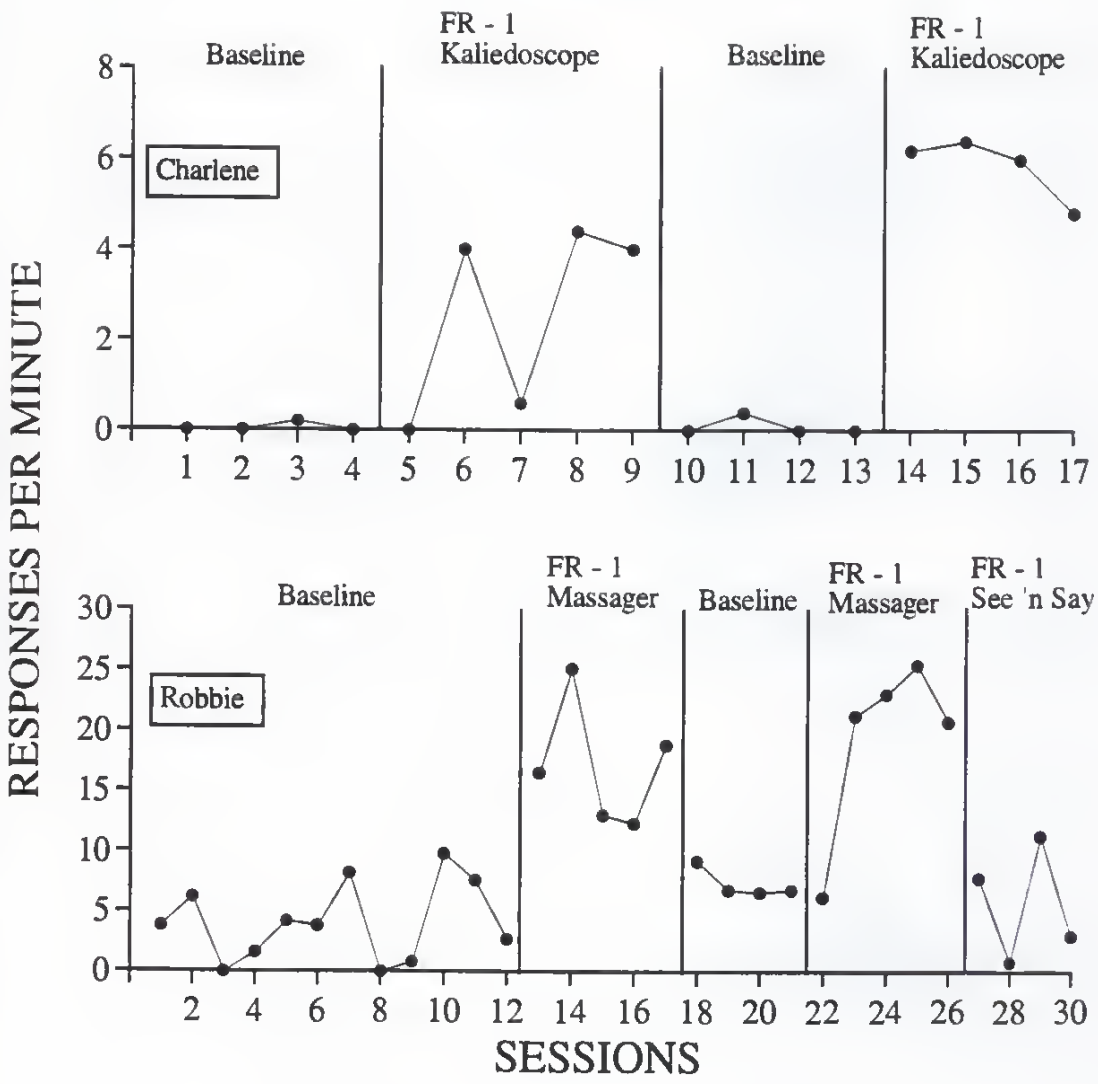


Figure 5. Responses per minute of adapative behavior for Charlene (top panel) and Robbie (bottom panel) during baseline and FR1 delivery of leisure items (Experiment 2).

duration engagement predicted that many of the items were reinforcers. Subsequent reinforcer assessments revealed that, when delivered on FR 1 schedules, two of the low selection/high duration items maintained levels of adaptive responding well above those observed in baseline, thereby validating the reinforcing efficacy of these items. A third item that failed to produce selection or duration percentages above 50%, also failed to increase levels of responding above baseline levels when delivered on an FR 1 schedule. As such, this provided a limited validation of the utility of the duration measure in predicting items that would not function as reinforcers.

It should be noted that the low selection percentages observed using the MSWO procedure may be less likely to be observed using a paired-choice procedure, depending on the number of highly preferred stimuli in the array. For example, in a paired-choice assessment using an array of 7 items, if two are highly and about equally preferred, one might expect selection of these two to be rather evenly divided when presented together as a pair, but that each would always be selected when paired with other items. Given five assessment sessions, this would result in 30 opportunities to select each of the two items. If one were chosen 28 times and the other 27 times, selection percentages of 93.3% and 90.0%, respectively would result. On the other hand, given the same conditions in a MSWO format, each item would have been chosen on the first or second trial during each assessment (thereafter both would have been unavailable), resulting in selection during five of the seven and five of the eight trials during which they were available. This would result in selection percentages of 71.4% and 62.5%. Using the same formulations with three highly preferred stimuli, the paired-choice procedure would produce percentages of 86.7%, 83.3%, and 80.0%, whereas the MSWO procedure would result in percentages of 55.%, 50.0%, and 45.5%. Thus, the paired-choice procedure produces higher selection percentages than does the MSWO procedure given a subset of items for which preference is roughly equivalent.

It was previously suggested that the duration measure might determine which of two factors might account for uniformly low selection percentages (i.e. that none of the stimuli was highly preferred or that many of the stimuli were highly preferred). However, none of the participants displayed patterns of duration engagement suggesting that all or most of the stimuli were low preference stimuli. That is, all engaged many of the items for high percentage durations. As such, it remains unknown if uniformly low selection percentages translate into low preference for all of the items. However, the failure of the See and Say toy in maintaining high responses rates for Robbie's coin matching are suggestive that low duration items are not effective reinforcers.

A related limitation of Experiment 2 was that, although the duration-based assessment suggested that many of the items were effective reinforcers, the effects of only one putative reinforcer were tested for each participant. The high duration engagement item tested for Robbie (the massager) was engaged for 90% of the time during the duration assessment. Four other items (book, pencil and paper, Connect 4, and puzzle) were engaged for over half the time they were available and it remains unknown if these items would maintain levels of adaptive responding above those observed in baseline. However, the item assessed for Charlene (kaleidoscope), was engaged for only 58% of the time it was available during the duration assessment. Four other items produced higher percentages of durations, suggesting that if engagement duration is an accurate index of reinforcement effects, these items would also have maintained high rates of adaptive responding. Therefore, although the results for Charlene suggest that items that are engaged for longer durations than they are not (i.e. 50% duration or more) might function as reinforcers, further evaluations are required to determine the extent to which percentages of item engagement directly translate into estimates of reinforcement efficacy. This could be done in a manner similar to that described by Piazza, Fisher, Hagopian et al. (1996) for the paired-choice assessment. That is, by comparing relative reinforcement effects between high-and middle duration items, and between middle- and

low-duration items, etc., one could determine the extent to which rankings based on engagement duration correlate with reinforcement effects when used in training situations.

Future research could also compare the results of approach and duration measurements using single-stimulus presentation formats for each type of assessment. It was previously suggested that the procedure described by Pace et al. (1985) might be prone to overestimation of reinforcement effects because participants may approach most or all of the items placed in front of them. The same limitation might apply to a duration-based assessment when presenting items individually. In this experiment, none of the items resulted in a 0% engagement duration, and only five of the items resulted in duration percentages of below 50%. It may be the case that participants are, in fact, unlikely to quickly release nonreinforcing items and that the items were picked up simply because they were the only item available. If so, we might not expect predictions based on duration to differ dramatically from predictions based on selection during single-stimulus presentation. Although not attempted in this experiment, this comparisons could be easily made by making both sorts of observations during a duration assessment. That is, in addition to measuring engagement durations, experimenters could note if the item was approached within the amount of time allotted during a single-stimulus approach assessment (e.g., within 5 s in the procedure described by Pace et al., 1985).

EXPERIMENT 3: EMERGENCE OF REINFORCER PREFERENCE AS A FUNCTION OF SCHEDULE REQUIREMENTS AND STIMULUS SIMILARITY

Procedurally, the paired-choice preference assessment (Fisher et al. 1992) can be described as series of concurrent FR 1 schedules of reinforcement. That is, two reinforcers are concurrently available, and the delivery of those reinforcers is contingent upon a single response (approach or selection). Tustin (1994) recently expressed concern that preference assessments in which the response requirement for delivery of reinforcement is low (e.g., FR 1 schedules) may not accurately predict preference under higher schedule requirements. He tested this possibility using concurrent schedules for which the ratio requirement changed across sessions. Specifically, the number of responses required to earn each of two available reinforcers changed simultaneously across sessions from values of FR 1 to FR 20. Tustin found that the participant's initial preference for one stimulus over another reversed as the schedule requirements increased; that is, the preference revealed under the concurrent FR 1 condition did not prove durable when the response requirement for obtaining reinforcement increased.

Tustin's (1994) data may have important implications for the selection of reinforcers used during training and treatment for individuals with developmental disabilities. To the extent that Tustin's results are generalizable, current assessment methods may make inaccurate predictions about reinforcer efficacy when the tasks used in training regimens require either more responses or more effort prior the delivery of reinforcement.

Experiment 3 is an attempt to replicate and extend Tustin's (1994) findings. His data on the effects of progressive concurrent schedules were based on observations of one individual (Subject #3), for whom only three sessions were conducted at each schedule value. As such, the extent to which similar results would be obtained with other individuals is unclear, and the possibility that observed findings were due to random

variability in a small sample of behavior cannot be ruled out. Experiment 3 represents an attempt to extend the generality of these findings by replicating the procedures with additional individuals through more extended phases at each schedule value. In addition, the reinforcers in Tustin's study, computer-generated visual patterns and tones, may have been somewhat atypical. In the present analysis, small edible reinforcers and manipulable items more commonly used as reinforcers in training or treatment programs were selected for assessment.

Experiment 3 also sought to initiate the investigation of conditions under which shifts in preference under identical but increasing schedules of reinforcement will and will not occur by making comparisons involving categorically similar versus categorically dissimilar reinforcers (i.e., those operating through the same sensory modality and those operating primarily through different sensory modalities). In Tustin's (1994) study, the reinforcers that were concurrently available overlapped to a certain degree. In the present study, the effects produced by increasing schedule requirements for similar and dissimilar reinforcers were examined by using two food items (similar reinforcers) and a food item versus a leisure item (dissimilar reinforcers) in separate analyses. It is possible that similar reinforcers may share more functional characteristics and show greater substitutability than do dissimilar reinforcers. If so, preference between similar reinforcers may be more sensitive to concurrent schedule manipulations than preference between dissimilar reinforcers.

Method

Participants, Setting and Materials

Two participants from Experiment 1, Eliza and Rudy, participated in the third experiment. Sessions were conducted in one room of a day treatment facility for individuals with behavior problems. Each room contained tables, chairs, and a small, wheeled cart that held the response apparatus. This apparatus consisted of a foam board on which two microswitch panels were mounted 4 cm apart. Each panel measured 20.5

cm by 12.5 cm. Depression of the panel resulted in the illumination of a small light in its center. Both panels were made of yellow plastic, but the left panel was covered with blue construction paper to aid in discriminability.

Seven stimuli were selected for use in the study based on the results obtained during a MSWO assessment procedure. Chocolate chip cookie halves, cheese crackers, and a vibrating massager were selected for Rudy. The experimenter turned on the massager prior to handing it to Rudy, who would invariably place it against his throat, providing stimulation to that area. The stimuli selected for Eliza included small pretzels, potato chips, an orange-flavored drink delivered in a small plastic cup, and an inflated balloon. When presented with the inflated balloon, Eliza would typically toss it in the air and bat it back and forth with the experimenter. In addition to other considerations described below, these items were selected because two sets were similar (e.g., cookie-cracker for Rudy, pretzel-potato chip for Eliza), but could be recombined to produce dissimilar sets (e.g., cookie-massager for Rudy, drink-balloon for Eliza).

Reinforcer Selection Procedure

Stimulus preference assessment. Several reinforcers were identified for both participants using the MSWO format (DeLeon & Iwata, 1996) described in Experiment 1. Two sets of assessments were conducted for each individual. The first included food items exclusively; the second included only leisure items. Thus, food and leisure items were assessed separately.

Pre-training. During this phase, participants were taught the contingency between switch pressing and delivery of preferred items. Both individuals had previous experience in the use of the microswitch and invariably pressed only one switch at a time throughout the study. However, the following three-phase procedure was implemented to ensure that they could discriminate pressing either one of the panels and delivery of the reinforcer associated with that panel. At the beginning of each session during the first phase, the cart containing the microswitch pads was rolled in front of the individual, who

was seated in a chair. A plate containing one of the previously identified food reinforcers was placed directly behind one of the panels. The experimenter then pointed to the panel that would produce reinforcement and asked individual to press that panel. If the individual pressed the correct panel with enough force to illuminate the small light, the reinforcer was delivered (FR 1 schedule). If the individual pressed the correct panel without enough force to illuminate the light, pressed the incorrect panel (which produced illumination throughout the study), or emitted no response, the experimenter physically prompted the correct response. On subsequent trials, the position of the plate was alternated randomly from the area behind one panel to the area behind the other panel. The same procedure was used during the second phase, except that the experimenter no longer pointed to the correct panel. During the third phase, an empty plate was added behind the inoperative panel, and the two plates were again placed randomly across trials. Ten trials were conducted during each session, and movement from one phase to the next occurred when the individual pressed the correct panel on 90% or more of the trials across three consecutive sessions.

Preference probes. Following completion of the pre-training sequence, probes were conducted to identify pairs of reinforcers for which preference was roughly equivalent. The comparisons were counterbalanced across individuals: Dissimilar reinforcers (foods versus leisure items) were tested for Rudy first, whereas similar reinforcers (two food items) were tested for Eliza first. The initial pair of reinforcers was selected based on similar rankings during the stimulus preference assessment and on their categorical similarity or dissimilarity (e.g., a similarly ranked food and a leisure item were first compared for Rudy). At the beginning of each session, the experimenter pushed the cart directly in front of the individual, placed the two reinforcers on plates (one behind each panel), and asked the individual to press either panel. Both panels were always operative throughout the session. The item located behind the pressed panel was delivered contingent upon each press (FR 1 schedule) and, unlike the training sessions, the

placement of items behind each panel remained the same throughout the session. Rudy was exposed to 24 trials during each session, whereas Eliza was exposed to 14 trials. If the individual allocated a much larger proportion of responses (e.g., two-thirds) toward the panel associated with one of the stimuli for two consecutive sessions, indicating noticeable preference for that reinforcer, the reinforcers were changed for the next session. This process continued until the individual's response allocation to a pair of reinforcers was roughly equivalent. If no differences or only small differences were noted in response allocation across two consecutive sessions, these two probe sessions became the first two sessions of the rest of the analysis, and the study proceeded as described below. The first set of both similar and dissimilar reinforcers evaluated for Rudy produced relatively equal preference between two items. Five sets of dissimilar reinforcers and 2 sets of similar reinforcers were evaluated for Eliza before meeting the equal preference criterion.

Effects of Increasing Schedule Requirements on Relative Preference

During all remaining phases, a response consisted of pressing either microswitch pad with enough force to illuminate the light in the center of the pad. Reinforcement was delivered according to concurrent FR schedules. Again, both panels were always operative and responses to either panel simply accumulated regardless of whether the participant switched between panels before the schedule requirement was met for either. When a schedule was completed on one of the panels, the therapist handed the individual the stimulus corresponding to the panel on which the required number of responses had been made. As in the pre-experimental probes, Rudy's sessions ended when 24 reinforcers had been delivered, and Eliza's ended when 14 reinforcers had been delivered (Note: Rudy's assessment was conducted first, and a number close to Tustin's [1994] was selected for the sake of replication. However, at the higher schedule values, session length was extremely long, sometimes upwards of 45 minutes. Therefore, all of Eliza's

assessments were based on either 14 trials [preference assessment and probes] or 14 reinforcers earned [schedule assessment]).

The schedules of reinforcement were increased progressively across phases. Both individuals began with concurrent FR 1 / FR 1 schedules. When responding (preference) stabilized, the ratio requirement for both reinforcers was simultaneously increased to FR 2, FR 5, FR 10, and in both cases of dissimilar reinforcers, FR 20. The schedule increases ended when no consistent preference was observed during the concurrent FR 20 / FR 20 comparison or at the value at which consistent preference was observed for one item over the other. At that point, a reversal to the original schedules was conducted for Eliza by returning to FR 1 schedules.

Response Measurement and Interobserver Agreement

Observers were graduate and undergraduate students who had demonstrated proficiency with the type of data collection used in this study by attaining a minimum of 90% agreement with a previously trained observer for three consecutive sessions. During the stimulus preference assessment, an observer recorded the order of selections. Items were subsequently ranked according to a percentage score calculated by dividing the number of times each item was selected across the five sessions by the number of times the item was available, and multiplying by 100%. During 20% of the trials, a second observer also recorded the order of selections. When comparing observers' records, an agreement was scored if both observers recorded the same order of selection for each item. Interobserver agreement was calculated by dividing the number of selections on which observers agreed by the total number of selections and multiplying by 100%. Across assessments, interobserver agreement never deviated from 100%.

During all sessions involving panel pressing, an observer recorded the number of responses on each panel and the number of each type of reinforcer delivered using a hand-held computer (Assistant, Model A102). Interobserver agreement was assessed by having a second observer record events during 39.3% of the sessions. When comparing

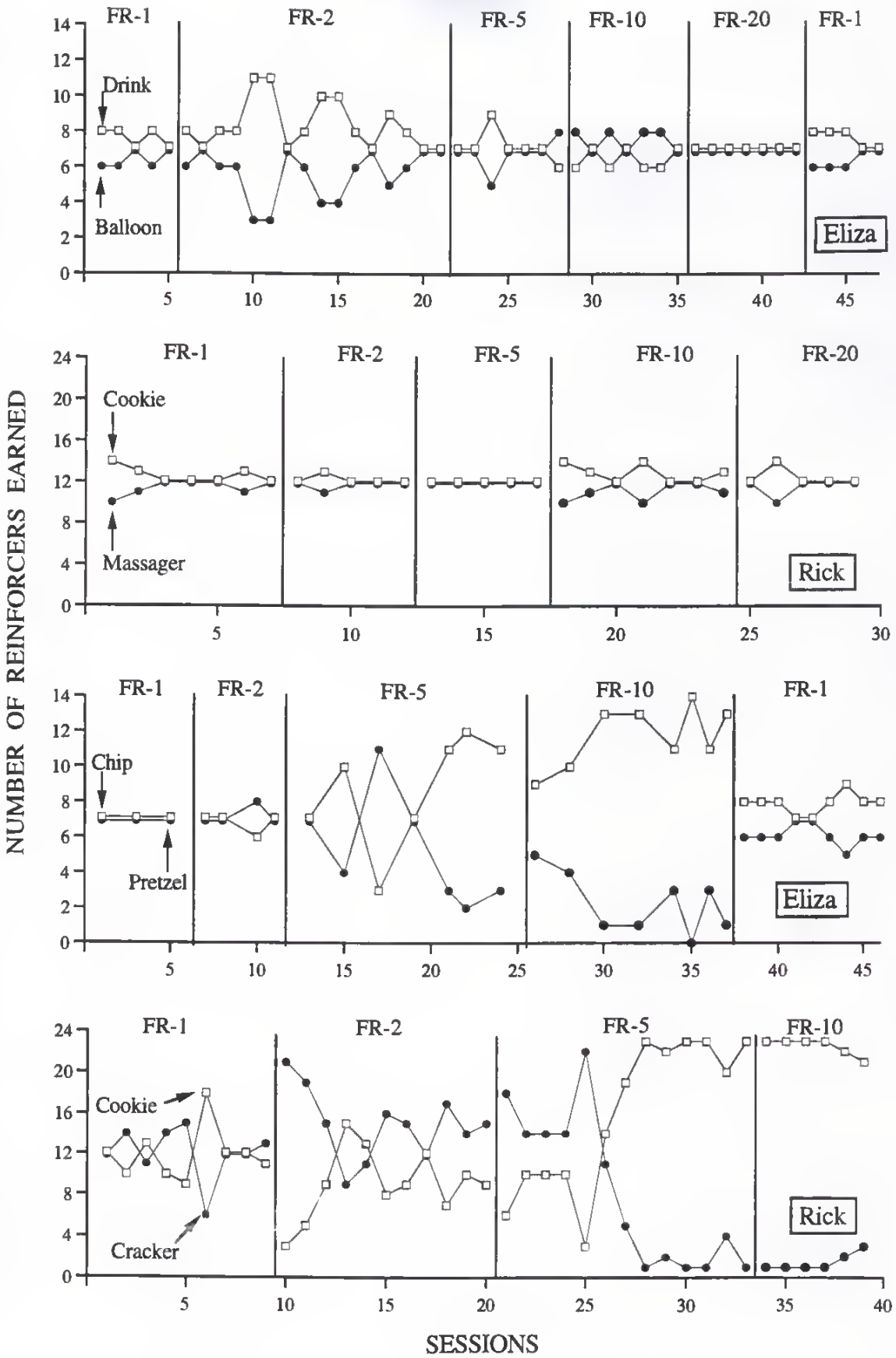
observers' records, session time was divided into 10-s intervals, and agreement was calculated on an interval-by-interval basis by dividing the smaller number of recorded events by the larger number. These quotients were then summed across intervals, divided by the total number of intervals in the session, and multiplied by 100%. Mean interobserver agreement for panel pressing was 96.3% (range, 86.3% to 100%). Mean agreement for reinforcer delivery was 98.3% (range, 90.7% to 100%).

Results

Figure 6 shows the allocation of responses by both participants during the comparison of dissimilar reinforcers (top two graphs) and similar reinforcers (bottom two graphs), expressed as the number of reinforcers earned per session from presses on the two panels. Data in the top graph show that when two dissimilar items (drink versus balloon) were concurrently available to Eliza on FR 1 schedules, responses were allocated slightly more towards the panel associated with the drink, although responding was roughly equivalent for both reinforcers. During the FR 2 condition, the drink was clearly more preferred during 5 of the 16 sessions, but no consistent preference was observed, in that choice allocation always reconverged toward an equal distribution. With the exception of Session 24 during the FR 5 condition, response allocation during the FR 5, FR 10, and FR 20 conditions revealed no consistent preference for either of the two stimuli. This lack of preference continued to be evident during the final condition, which consisted of a return to concurrent FR 1 schedules.

A similar pattern was observed for Rudy, whose data in the second graph show results obtained during the comparison of dissimilar reinforcers (cookie versus massager). Rudy showed little preference between the cookie and massager when both were concurrently available on FR 1, and subsequently on FR 2 and FR 5 schedules of reinforcement. At FR 10, Rudy seemed to show a slight preference for the cookie, but it later disappeared when both reinforcers were available on an FR 20 schedule.

Figure 6. Response allocation for dissimilar (top two panels) and similar (bottom two panels) reinforcers by both participants during concurrent fixed ratio (FR) schedules, expressed as the number of reinforcers earned under each schedule (Experiment 3).


FR-1
FR-2
FR-5
FR-10
FR-20
FR-1
Drink
Balloon
Eliza
Cookie
Massager
Rick
Chip
Pretzel
Eliza
Cookie
Cracker
Rick
NUMBER OF REINFORCERS EARNED
SESSIONS

During Eliza's comparison of similar reinforcers (chip versus pretzel, third graph), initial daily inspection of the data revealed systematic differences in her allocation of choices, which were related to temporal sequence, beginning with the FR 5 comparison. That is, when two sessions were conducted on the same day, the first session consistently showed either no clear preference or a preference for the chip, whereas the second session of the day showed a preference for the pretzel. This consistent shift in preference was attributed to satiation to the chips during the first session. Subsequently, the entire analysis up to 33rd session was replotted using only the first session of the day and, beginning with the 34th session, only one session was conducted per day. This analysis revealed no consistent preference during the FR 1 and FR 2 conditions. During the FR 5 condition, some between-session switching in preference was observed initially, but a clear preference for the chip emerged over the last 3 sessions, and this trend continued throughout the FR 10 condition. When the schedule requirements were reset to FR 1, preference that had emerged through the FR 5 and FR 10 conditions was greatly diminished such that the distribution of choices appeared roughly similar to that observed during the initial FR 1 comparison.

Data in the bottom graph show results obtained during the comparison of similar reinforcers (cookie versus cracker) for Rudy. At FR 1, there was a great deal of between-session switching with respect to the item chosen most frequently. At the beginning of the FR 2 condition, Rudy chose the cracker more often, but this pattern switched on Session 13 and then reemerged to a lesser extent from Session 15 onward. Preference for the cracker continued during the first few sessions of the FR 5 condition, but switched on Session 26. Thereafter, a notable preference for the cookie was observed for the remainder of the FR 5 condition and throughout the FR 10 condition.

Discussion

When two similar reinforcers (foods), for which preference was established to be roughly equivalent, were available under concurrent FR 1 schedules, both participants

showed little preference for either stimulus. However, increasing the schedule requirements revealed clear preference by both individuals for one of the foods during the FR 5 condition, and these preferences maintained when the schedules were again increased to FR 10. Thus, the findings with respect to similar reinforcers are comparable to those reported by Tustin (1994). By contrast, neither individual showed a strong preference for one of the dissimilar reinforcers (a food versus a leisure item) across progressive schedule increases, even at the FR 20 value. These results differed markedly from those reported by Tustin: His participant showed a modest switch in preference under FR 2 schedules, pronounced preference at the FR 10 value, and exclusive responding for the preferred reinforcer at the FR 20 value. However, it must be noted that shifts in preference may be idiosyncratic across individuals. It is possible that schedule requirements higher than FR 20 may have produced shifts in preference for dissimilar reinforcers by participants in Experiment 3, although such schedules may not be encountered frequently under typical training conditions.

Clear and consistent preferences emerged when initially equivalent and similar reinforcers (foods) were compared under increasing schedule requirements. This finding suggests that, for some classes of reinforcers, simultaneous increases in price may magnify small differences in preference that are undetected when work requirements are low. In such cases, a stimulus preference assessment involving low response requirements (FR 1 schedules) would not accurately predict the relative potency of reinforcers under increased response requirements.

Because Rudy was exposed only to increasing schedule values across phases, with no reversal to a previous value, it is possible that emergent preference between similar reinforcers may be partially a function of the passage of time or sequence effects. However, Eliza's preference for chips, which emerged under FR 5 and FR 10 requirements, largely disappeared when the schedules were reset to FR 1. Similarly, Tustin (1994, see Subject 3) used increasing, decreasing, and then increasing schedule

values across three conditions and observed a clear reversal between the preference patterns observed under low versus high schedule requirements. Taken together, the data for Eliza and those presented by Tustin suggest that shifts in preference under high schedule requirements may be reversible and not due solely to temporal or sequence effects.

The different outcomes observed when similar versus dissimilar reinforcers were compared must be considered preliminary because data were presented for only 2 participants. Furthermore, Tustin (1994) observed a reversal of preference between two stimuli that might be considered dissimilar in several respects. Nevertheless, to the extent that these results are reliable, they may reflect the effects of increasing price requirements on choice between reinforcers that are substitutable to a greater or lesser degree. The relationship between substitutable reinforcers is such that an increase in the price of one reinforcer (e.g., an increase in the ratio requirement) leads to decreased consumption of that reinforcer and increased consumption of a concurrently available substitute (Hursb, 1980). For example, to the extent that Coke® and Pepsi® are substitutable, increases in the price of Coke® should produce decreases in its consumption and corresponding increases in the consumption of Pepsi®. Green and Freed (1993) suggested that substitutable stimuli are frequently those that serve similar functions (e.g., Coke® and Pepsi® both alleviate thirst and have a cola taste), and that consumption of a particular reinforcer is influenced by the availability of other reinforcers serving the same function.

Continuation of this analogy might help to explain the results obtained in the present study. Assuming that Coke® and Pepsi® are both available for a dollar per serving and that a person has only a slight preference for Coke®, the individual may allocate choices rather evenly, perhaps as a function of periodic satiation for the preferred item, but with slightly more overall selections of Coke®. Now assume that the cost of each is increased to $5 per serving. At this price, the preference for Coke® is likely to be expressed. By contrast, a similar arrangement involving Coke® and bus tokens may

produce different results. Again, at a dollar per item, roughly equal selection between the two options would not be surprising, assuming that the establishing operation for each dictates that both are momentarily equally valuable. However, these items serve distinctly different functions and are not substitutable; that is, the person is not free to trade one for the other and continue to receive functionally similar reinforcement at the same rate. The person is more likely to continue choosing equally, even when the price for both reinforcers increases substantially.

The same might be said for results obtained in the present study. When choices involved two substitutable items, such as a cookie and a cracker, concurrent increases in the cost of each might may have "forced" the expression of slight preference for one of the items. However, when reinforcers that were unlikely to be substitutes, such as a cookie and a massager, were concurrently available and equally preferred, increases in cost had little effect on preference.

It is possible that these differences were also a function of the experimental arrangement. Given that the comparisons were made under a concurrent arrangement (in which responding for one reinforcer foregoes delivery of the other) with a fixed number of reinforcers available, the individual had little to lose by foregoing the less preferred reinforcer when the options were similar because food was obtained in either case, but not when the reinforcers were dissimilar. It is not clear if these patterns would be observed under an arrangement in which each reinforcer is presented singly in different conditions (i.e., under separate conditions consisting of either low or high FR values). Change of preference in such a context would be reflected in different response rates associated with the different reinforcers under low versus high schedule values, and, as noted previously, values higher than those used in this study might have produced a change in preference for dissimilar reinforcers.

Another difference between data in Experiment 3 (for similar reinforcers) and those reported by Tustin (1994) is that I observed the emergence of preference for one of two

reinforcers that were initially preferred equally, whereas Tustin observed a reversal of preference (i.e., the more preferred reinforcer under the low schedule value became the less preferred reinforcer under the high schedule value). Although these may represent different behavioral effects whose relative prevalence is unknown, either outcome suggests that preference (or lack thereof) may be altered as a function of increasing schedule requirements. It would be worthwhile in future studies to determine how common such results are. If the emergence or reversal of preference occurs infrequently, assessment methods incorporating low response requirements can continue to be used with little concern. However, if the phenomena are commonly observed, reinforcer identification methods may yield more accurate information if response requirements during assessment approximate those used in the eventual training context.

Finally, the present data may also suggest practical strategies with respect to reinforcer usage in different situations. For example, it appeared that cookies and crackers may be equally effective reinforcers for Rudy under low work requirements, but that cookies may be more effective under high requirements. If so, it may prove beneficial to reserve cookies for particularly effortful tasks and to restrict their usage during less effortful tasks, given that crackers may work just as well under the latter conditions. However, a more thorough analysis is required before it can be concluded that differences of the type found here actually translate to differences for everyday tasks in which relative response effort cannot be measured in units as simple and discriminable as panel presses. Furthermore, even the less preferred reinforcers maintained some responding during the higher ratio values, and it is likely that, outside of a comparison of relative preference, both reinforcers might maintain similar rates of responding if presented as the only reinforcement option (Lerman et al., in press; Smith et al., 1995). Future research might apply the type of analysis used here with tasks more typical of those used during training to determine if differences in relative preference generalize to differences in rate or accuracy of responding with more socially significant activities. If

such differences are borne out, briefer versions of this analysis, perhaps similar to that used by Tustin (1994), in which only a few comparisons are made at each schedule value, may prove beneficial in maximizing reinforcement effects during training.

GENERAL CONCLUSIONS

In this series of investigations, it was observed that several factors can influence the outcome of preference assessments. The availability of multiple food items in the selection array was shown to mask preferences for effective nonfood reinforcers. The presence of several high-preference items (as determined through a duration-based assessment) resulted in low selection percentages for all items during selection-based assessments. Finally, under some conditions, relative preferences between items under low schedule requirements did not correspond to relative preferences when schedule requirements were considerably higher.

One implication of these findings is that different procedural variations of preference assessments may be necessary depending upon the task at hand and upon initial results. As previously mentioned, if caregivers wish to identify nonfood reinforcers, then those stimuli may have to be assessed in the absence of food. Furthermore, duration-based assessment may be necessary when initial assessments produce uniformly low selection percentages. Arguably, one could eliminate this two-step process simply by omitting the selection-based procedure. However, by their very nature, duration-based assessments are time-consuming and perhaps impractical for daily implementation just prior to training or treatment sessions. Therefore, if the aim is to select reinforcers that will be used consistently over the course of training or treatment, duration-based assessments may be the preferred option in that they may be less likely to produce ambiguous results. However, if the aim is to find those reinforcers that are momentarily the most potent, a selection-based approach will more readily facilitate frequent preference assessments. Finally, assessments that involve low response

requirements to indicate preferences may be adequate if the task to be trained is also of relatively low effort. However, if the target response is highly effortful, then it may be best to assess preferences under similar conditions. Perhaps previously established, high-effort behavior could be used as a selection response. Again, this adds considerable time and effort to the assessment, but perhaps reduces the possibility that treatment will fail due to ineffective reinforcement contingencies and, in the long, may eliminate the need to restart treatment attempts.

The same sort of considerations apply to methods of evaluating reinforcement effects. That is, reinforcer assessment methods should also be selected with consideration for the goals of the assessment. The first two experiments both used a single-operant arrangement to test reinforcement effects. Specifically, reinforcement effects were determined by comparing rates of responding observed when a single item was made contingent upon correct responses with rates observed during baselines in which only verbal praise was delivered for correct responding. The third experiment, however, used concurrent ratio schedules to determine relative reinforcement effects. The single- and concurrent-operant procedures have relative advantages and disadvantages that make each better suited for slightly different purposes. Fisher and colleagues (Fisher et al., 1996; Fisher, Thompson, Piazza, Crosland, & Gotjen, in press) have argued that concurrent operant arrangements are more sensitive to relative reinforcer effects than single-operant arrangements. That is, given two reinforcers of slightly different value, each might produce a similarly high rate of responding if presented contingently in a single-operant arrangement. However, in a concurrent-operant arrangement, one would expect the individual to differentially allocate responding towards the option associated with the more valued stimulus. For example, one method of manipulating reinforcer preference is to differ the response requirement for each of two identical and concurrently available reinforcers such that a preference is shown for the reinforcer associated with the lower schedule requirement. Basic research on concurrent ratio schedules arranged to

deliver reinforcers that differ along this dimension have shown near exclusive responding for the reinforcer associated with the higher rate of reinforcement (Herrnstein & Loveland, 1975) . As such, the concurrent-operant arrangement may be better suited for determining the more valuable of two reinforcers relative to each other.

However, exclusive responding towards the option associated with the higher value provides information about relative reinforcer value, but does not address the value of the stimulus that is never selected. This is a problem similar to that associated with the group presentation method described by Windsor et al. (1994) and the duration assessment described by Ringdahl et al. (1997). That is, if the individual selects one stimulus to the total exclusion of the other, no information is gained regarding the absolute value of the unselected stimulus. It may be that case that if the unselected stimulus was assessed individually, it would be shown to be a very potent reinforcer, but it's reinforcement value was masked by the availability of a more valuable item. As such, the single operant arrangement might be better suited in the determination of absolute reinforcer value (i.e., to determine simply if the stimulus does or does not function as a reinforcer).

These are just a few of the circumstances under which different experimental arrangements might be better suited for answering different sorts of questions regarding reinforcer efficacy. Certainly, much remains to be learned about the relative effects of reinforcers as well as about experimental preparations that can answer related questions. The basic learning literature includes many investigations of variables that affect choices among concurrently available reinforcers. A large proportion of these studies involve the matching law, which states that the distribution of responding under concurrent schedules of reinforcement will approximate the rates of reinforcement available under those schedules (Herrnstein, 1961). Recent work in applied behavior analysis has followed suit by beginning to examine variables that influence choices among reinforcement options. Much of this research has involved extensions to human behavior of laboratory studies on the matching law. For example, Mace et al. (1990) showed that when a student received

reinforcement for completing either multiplication or division problems on concurrent variable ratio 2 (VR 2) schedules of food delivery, the student completed roughly equal numbers of the two types of problems. However, when the rate of reinforcement for multiplication problems was doubled by lowering its schedule requirement to FR 1, the student completed far more multiplication problems than division problems.

This study and others like it serve as important demonstrations of human matching and illustrate the effects of varying reinforcement parameters, such as rate, quantity, or delay, on choice. However, the type of reinforcement delivered in such studies has typically been held constant. That is, unlike preference and reinforcer assessments, the same class of reinforcers are made available. Thus, relationships that might be observed when categorically different reinforcers are delivered are less clear (Hursh, 1980; Neef, Mace, Shea, & Shade, 1992). This is an important topic for consideration because, in many applied situations, choice often involves selection from among reinforcers that vary along numerous dimensions (Fuqua, 1984). Areas of basic research that have investigated relationships between concurrently available and categorically distinct reinforcers, although few in number, have been very influential in providing a background for work on preference and reinforcer assessment. For instance, duration-based assessments are loosely based on the Premack principle (Premack, 1962) and concepts from the field of behavioral economics (Green & Freed, 1993; Hursh 1980) have been used to account for observations regarding choices between reinforcers for individuals with developmental disabilities (Shore et al., 1997; Tustin, 1994). But again, lines of basic research that explore relationships between different types of reinforcers are probably the exceptions. Thus, in addition to providing methods to improve training and treatment programs for individuals with developmental disabilities, continued research on preference and reinforcer assessments may provide useful tools for the basic understanding of relationships among categorically distinct reinforcers.

REFERENCES

Allyon, T. & Azrin, N. H. (1968). Reinforcer sampling: A technique for increasing the behavior of mental patients. Journal of Applied Behavior Analysis, 1, 13-20.

Belfiore, P. J., Lee, D. L., Vargas, A. U., & Skinner, C. H. (1997). Effects of high-preference single-digit mathematics problem completion on multiple-digit mathematics problem performance. Journal of Applied Behavior Analysis, 30, 327-330.

Cautela, J. R., & Kastenbaum, R. A. (1967). A reinforcement survey for use in therapy, training, and research. Psychological Reports, 20, 1115-1130.

Corte, H. E., Wolf, M. M., & Locke, B. J. (1971). A comparison of procedures for elimination of self-injurious behavior of retarded adolescents. Journal of Applied Behavior Analysis, 4, 201-213.

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. Journal of Applied Behavior Analysis, 29, 519-533.

Egel, A. L. (1981). Reinforcer variation: Implications for motivating developmentally disabled children. Journal of Applied Behavior Analysis, 14, 345-350.

Fisher, W. W., Piazza, C. C., Bowman, L. G., & Amari, A. (1996). Integrating caregiver report with a systematic choice assessment. American Journal on Mental Retardation, 101, 15-25.

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. Journal of Applied Behavior Analysis, 25, 491-498.

Fisher, W. W., Thompson, R. H., Piazza, C. C., Crosland, K., & Gotjen, D. (in press). On the relative reinforcing effects of choice and differential consequences. Journal of Applied Behavior Analysis.

Fuqua, R. W. (1984). Comments on the applied relevance of the matching law. Journal of Applied Behavior Analysis, 17, 381-386.

Green, C. W., Reid, Canipe, V. A., & Gardner, S. M. (1991). A comprehensive evaluation of reinforcer identification processes for persons with profound multiple handicaps. Journal of Applied Behavior Analysis, 24, 537-553.

Green, C. W., Reid, D. H., White, L. K., Halford, R. C., Brittain, D. P., & Gardner, S. M. (1988). Identifying reinforcers for persons with profound handicaps: Staff opinion versus systematic assessment of preferences. Journal of Applied Behavior Analysis, 21, 31-43.

Green, L., & Freed, D. E. (1993). The substitutability of reinforcers. Journal of the Experimental Analysis of Behavior, 60, 141-158.

Herrnstein, R. J. (1961). Relative and absolute strength of a response as a function of frequency of reinforcement. Journal of the Experimental Analysis of Behavior, 4, 267-272.

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. Journal of the Experimental Analysis of Behavior, 24, 107-116.

Hursh, S. R. (1980). Economic concepts for the analysis of behavior. Journal of the Experimental Analysis of Behavior, 34, 219-238.

Ivancic, M. T. & Bailey, J. S. (1996). Current limits to reinforcer identification for some persons with profound multiple disabilities. Research in Developmental Disabilities, 17, 77-92.

Kazdin, A. E. (1984). Behavior modification in applied settings. Homewood, IL: The Dorsey Press.

Lerman, D. C., Iwata, B. A., Rainville, B., Adelinas, J., Crossland, K., & Kogan, J. (in press). Effects of reinforcement choice on task responding in individuals with developmental disabilities. Journal of Applied Behavior Analysis.

Mace, F. C., McCurdy, B., & Quigley, E. A. (1990). A collateral effect of reward predicted by matching law. Journal of Applied Behavior Analysis, 23, 197-205.

Mason, S. A. , McGee, G. G., Farmer-Dougan, V., & Risley, T. R. (1989). A practical strategy for ongoing reinforcer assessment. Journal of Applied Behavior Analysis, 22, 171-179.

Michael, J. (1982). Distinguishing between discriminative and motivational functions of stimuli. Journal of the Experimental Analysis of Behavior, 22, 171-179.

Neef, N. A., Mace, F. C., Shea, M. C., & Shade, D. (1992). Effects of reinforcer rate and reinforcer quality on time allocation: Extensions of matching theory to educational settings. Journal of Applied Behavior Analysis, 25, 691-699.

Pace, G. M., Ivancic, M. T., Edwards, G. L., Iwata, B. L., & Page, T. A. (1985). Assessment of stimulus preference and reinforcer value with profoundly retarded individuals. Journal of Applied Behavior Analysis, 18, 249-255.

Paclawskyj, T. R., & Vollmer, T. R. (1995). Reinforcer assessment for children with developmental disabilities and visual impairments. Journal of Applied Behavior Analysis, 28, 219-224.

Piazza, C. C., Fisher, W. W., Hagopian, L. P., Bowman, L. G., & Toole, L. (1996). Using a choice assessment to predict reinforcer effectiveness. Journal of Applied Behavior Analysis, 29, 1-9.

Piazza, C. C., Fisher, W. W., Hanley, G. P., Hilker, K., & Derby, K. M (1996). A preliminary procedure for predicting the positive and negative effects of reinforcement-based procedures. Journal of Applied Behavior Analysis, 29, 137-152.

Premack, D. (1962). Reversibility of the reinforcement relation. Science, 136, 235-237.

Quilitch, H. R., Christopherson, E. R., & Risley, T. R. (1977). The evaluation of children's play materials. Journal of Applied Behavior Analysis, 10, 501-502.

Rincover, A., & Newsom, C. D. (1985). The relative motivational properties of sensory and edible reinforcers in teaching autistic children. Journal of Applied Behavior Analysis, 18, 237-248.

Rincover, A., Newsom, C. D., Lovaas, O. I., & Koegel, R. L. (1977). Some motivational properties of sensory stimulation in psychotic children. Journal of Experimental Child Psychology, 24, 312-323.

Ringdahl, J. E., Vollmer, T. R., Marcus, B. A., & Roane, H. S. (1997). An analogue evaluation of environmental enrichment: The role of stimulus preference. Journal of Applied Behavior Analysis, 30, 203-216.

Shore, B. A., Iwata, B. A., DeLeon, I. G., Kahng, S. & Smith, R. G. (1997). An analysis of reinforcer substitutability using object manipulation and self-injury as competing responses. Journal of Applied Behavior Analysis, 30, 21-41.

Singh, N. N., & Millichamp, C. J. (1987). Independent and social play among profoundly mentally retarded adults: Training, maintenance, generalization, and long-term follow-up. Journal of Applied Behavior Analysis, 20, 23-34.

Smith, R. G., Iwata, B. A., & Shore, B. A. (1995). Effects of subject- versus experimenter-selected reinforcers on the behavior of individuals with profound developmental disabilities. Journal of Applied Behavior Analysis, 28, 61-71.

Tustin, R. D. (1994). Preference for reinforcers under varying schedule arrangements: A behavioral economic analysis. Journal of Applied Behavior Analysis, 27, 597-606.

Vollmer, T. R., & Iwata, B. A. (1991). Establishing operations and reinforcement effects. Journal of Applied Behavior Analysis, 24, 279-291.

Vollmer, T. R., Marcus, B. A., & LeBlanc, L. (1994). Treatment of self-injury and hand mouthing following inconclusive functional analysis. Journal of Applied Behavior Analysis, 27, 331-344.

Windsor, J., Piche, L. M., & Locke, P. A. (1994). Preference testing: A comparison of two presentation methods. Research in Developmental Disabilities, 15, 439-455.

BIOGRAPHICAL SKETCH

Iser Guillermo DeLeon was born in Camaguey, Cuba, and emigrated to the United States in 1968. After several years in Chicago, Illinois, he and his family moved to Florida. Here, he attended the University of Florida and graduated with a degree in psychology in 1988. Following graduation, he found employment in Gainesville, first providing behavior therapy services for individuals with developmental disabilities and, then, supervising a small team of behavior management specialists. His experiences in this field led him to seek graduate training in the field of behavior analysis, and he subsequently enrolled in the applied behavior analysis program at Western Michigan University. Upon receiving a master's degree under the supervision of Dr. Wayne Fuqua, he returned to the University of Florida to seek a doctoral degree. For the past four years he has been working as a graduate research assistant at the Florida Center on Self-Injury, again involved in the treatment of behavior disorders, under the supervision of Dr. Brian Iwata. Expecting graduation in the Fall of 1997, Mr. DeLeon is about to begin a post-doctoral fellowship in the department of Pediatrics at the Johns Hopkins University School of Medicine, where he will continue clinical work and research in the application of behavior analytic principles to the treatment of behavior disorders.

I certify that I have read his study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Brian A. Iwata, Chair
Professor of Psychology

I certify that I have read his study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Marc N. Branch
Professor of Psychology

I certify that I have read his study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mary K. Dykes
Professor of Special Education

I certify that I have read his study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mark H. Lewis
Professor of Neuroscience

I certify that I have read his study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Timothy D. Hackenberg
Associate Professor of Psychology

This dissertation was submitted to the Graduate Faculty of the Department of Psychology in the College of Liberal Arts and Sciences and to the Graduate School was accepted as partial fulfillment of the requirement for the degree of Doctor of Philosophy.

December 1997

Dean, Graduate School